Spectroscopic studies of the biotin biosynthase enzymes

Nick Tomczyk

PhD

The University of Edinburgh

2006

Declaration

I, Nick Tomczyk, hereby certify that this thesis has been composed by myself, that it is a record of my work, and that it has not been accepted in partial or complete fulfilment of any other degree or professional qualification.

Nick Tomczyk
University of Edinburgh
2006

Acknowledgements

Firstly, I would like to acknowledge Dr Dominic Campopiano for the opportunity to carry out this study and I would like to thank him for his continued support and supervision on both a personal and professional level. I would also like to thank Prof Bob Baxter for his constant guidance, encouragement and advice during my period in the department.

I would also like to take the opportunity to thank all friends and colleagues from the Chemistry department as a whole and the Baxter and Campopiano groups. From Lab 229 a particular mention should go to, in no particular order, Lisa McIver, Sander Henzing, Dave Clarke, Rachel Breen, Mhairi Brunton, Ross Langley and all others for their continued helpful input and diversionary tactics alike….

I would also like to add a special thanks to Jen Lynch for her own 'unique' help, support and friendship. A big shout out to Perdita Barran and Nick Polfer for much scientific input, as well as much not so science orientated support.

Thanks also to Professor Marc Fontecave, Sandrine Ollagnier-de-Choudens, Jon Rubach and others from this group in Grenoble for providing many BS samples and for their constant input into the BS-PLP work.

Without the following this work would not have been completed and I offer them my thanks. Andy Cronshaw, Sally Shirran, Nathan Harris, Hannah Florence,

Finally I particularly thank Jo, and Mum, Dad and all the members of my family for all the constant support and encouragement throughout my studies and believing that I someday may indeed stop being a student and enter the world of 'working for a living'.

Abstract

Biotin synthase (BS) catalyses the final sulfur incorporation step in the biosynthesis of biotin. Its chemical mechanism, primary amino acid sequence and tertiary structure place it in the Radical-SAM-dependent superfamily of >600 proteins. The gene encoding *E. coli* biotin synthase (*bio*B) has been expressed as a histidine-fusion protein (6HisBS) and the recombinant protein purified in a single step using immobilised metal-affinity chromatography (IMAC). In addition, a number of single and double amino acid mutations of the conserved cysteine residues within the 'cys box' were constructed and purified using similar methodology. Wild-type and mutant 6HisBS proteins were compared using mass spectrometry (LC-ESI-MS) and UV-visible spectroscopy to probe characteristics altered by the mutation.

Biotin synthase was analysed for its ability to bind the cofactor pyridoxal 5'-phosphate (PLP). Two single point mutations (K49Q and K49R) of a putative PLP-binding residue, Lys 49, were isolated and compared with wild-type BS using a combination of biochemical and mass spectrometry techniques. Each protein was subjected to enzymatic proteolysis and peptide modification analysed by means of MALDI-ToF MS, ESI-QTOF MS and MS/MS. We identified a unique Lys49-containing fragment of the 6HisBS which corresponds to the PLP modified peptide. In contrast, this modified peptide is absent in the mutant proteins. This observation is further confirmed by MS/MS sequencing of this MS observed species which shows CID of both the peptide backbone and covalently attached PLP moiety.

Biotin synthase was prepared in the presence of excess PLP and its interaction with cyteine monitored with UV-visible spectroscopy. We observed marked changes in the UV-vis profile of biotin synthase under these conditions which is consistent observations of PLP-dependent cysteine desulferase enzymes.

Isolation of overexpressed biotin synthase yields protein preparations with variable rations of iron sulfur clusters. To study changes in global protein expression ratios during biotin synthase overexpression, comparative, 2-dimensional SDS-PAGE experiments were conducted. Selected visualised protein spots were identified via enzymatic

digestion and mass spectrometric analysis. No significant protein up-regulation of other *E. coli* proteins was observed.

The X-ray crystal structure of *E. coli* biotin synthase was published during this work. Data from a single crystal was obtained with protein re-loaded with iron and sulfur, thus providing a single snapshot of the structure. Biotin synthase and other Radical-SAM proteins have a $(\beta/\alpha)_6$ structural core. To further probe the dynamics of the BS protein a series of techniques were used to compare the wild-type 6HisBS and mutant proteins. Both high resolving gel filtration and cation exchange chromatography were used to assess the heterogeneity of the purified proteins. Three clearly-distinct species were observed in both mobility and isoelectric point (pI). These three species are affected to various degrees with the mutation of one or more conserved residues suggesting major structural roles. Purified proteins were subjected to time-dependent, limited enzymatic digestion and SDS-PAGE visualisation. Protein fragments were mapped to amino acid sequence using a secondary, in gel, digestion and MALDI-TOF analysis. This confirmed the solvent-exposed regions of the protein identified in the crystal structure and identified an enzymatically-labile site. This site is observed at a loop region, and proteolysis splits the protein monomer structure into two pseudo-domains.

To confirm this sub-domain structure, the *bio*B DNA gene sequence was used to construct two *bio*B "pseudo-genes" (encoding N- and C-terminal fragments) which were expressed in *E. coli*. The proteins expressed from each pseudo-gene were insoluble when produced separately, however duel expression of both genes yielded a significant quantity of soluble protein which displayed many of the physical characteristics of intact biotin synthase. These sections remained intact upon LC-MS analysis, which would suggest that each sub-domain has folded in vivo and formed a stable complex. These studies highlight the possibility of further sub-structures within the recognised Radical SAM $(\beta/\alpha)_6$ scaffold.

Abbreviations

AB – ammonium bicarbonate buffer

ACP – acyl carrier protein

AcN - acetonitrile

Ado - adenosine

AgCl – silver chloride

AMP – adenosine monophosphate

AON – 8-amino 7-oxononanoate

AONS – 8-amino 7-oxononanoate synthase

ARR – anaerobic ribonucleotide reductase

ARR-AE – anaerobic ribonucleotide reductase activating enzyme

ATP – adenosine 5'-triphosphate

AUC – analytical ultracentrifugation

BCCP – biotin carboxyl carrier protein

bp - base pair

BS – biotin synthase

BS_MF – biotin synthase provided by Prof. Marc Fontecave

CD – circular dichroism

CO_2 – carbon dioxide

CoA – coenzyme A

CV - Column volume

Cys - Cysteine

Da – Daltons

DAN – 7,8-diaminononanoate

DANS – 7,8-diaminononanoate synthase

DNA – deoxyribonucleic acid

DOA – deoxyadenosyl

DOA$^{\bullet}$ - deoxyadenosy radical

DTB – dethiobiotin

DTBS – dethiobiotin synthase

DTT – Dithiothreitol

E_1' – oxidised / semiquinone redox couple

E_2' – semiquinone / hydroquinone redox couple

EDTA – ethylene diaminetetracetic acid

ESI - electrospray

FAD – flavin adenine dinucleotide

Fe - iron

FLAV - *E. coli* Flavodoxin

FRED - *E. coli* Flavodoxin (ferredoxin) $NADP^+$ oxidoreductase

FMN – flavin mononucleotide

GdnHCl – Guanidine hydrochloride

HCl – hydrochloric acid

His_6 - six consecutive histidine residues

HPLC – high pressure liquid chromatography

hq - hydroquinone

ICP-AES - Inductively coupled plasma - atomic emission spectroscopy

IMAC - Immobilised metal affinity chromatography

IPTG - Isopropyl - β, D - thiogalactopyranoside

ISC – Iron sulfur cluster

LAM - lysine 2,3-aminomutase

LB – Luria Bertani medium

MALDI – matrix assisted lazer desorption ionisation

Mg - magnesium

NADPH – nicotinamide adenine dinucleotide phosphate

ox - oxidised

P450 – cytochrome P450 reductase

PCR – polymerase chain reaction

PEG – polyethylene glycol

PFL - pyruvate formate lyase

PFL-AE - pyruvate formate lyase activating enzyme

PLP – pyridoxal phosphate

Pt – platinum

RNA – ribonucleic acid

SAM - S -Adenosyl methionine

SDS PAGE - sodium dodecyl sulphate polyacrylamide gel electrophoresis

SE – sedimentation equilibrium

sq – semiquinone

SV – sedimentation velocity

TAE – Tris acetate EDTA

TB – transformation buffer

TCA - Trichloroacetic acid

TFA - Trifluoroacetic acid

ToF – time of flight

Tris – Tris [hydroxymethyl] aminomethane

tRNA – transfer ribonucleic acid

Q-ToF – Quadrupole time of flight

Table of Contents

List of Figures

Chapter One: Introduction

Chapter Three: Results and Disscussion

Chapter One: Introduction

1.1 – Introduction to biotin

The first recorded work on what was to become known as biotin took place in 1901 where a substance was isolated from beer wort by Wilders which he discovered stimulated the growth of yeasts. He termed this substance *bios* (4) but little work was done in the following 40 years until Williams found that the substance termed "bios" was in fact a mixture of biotin, alanine, myoinositol and pantothenic acid (17). At this time "protective compound X" and "vitamin H" were isolated from mammalian livers, and both were identical to biotin and were found to protect rats from ingestion of dried egg whites. Kögl had isolated pure biotin in 1936 from dried egg yolks as its methyl ester (18) and it has since been found to be an essential vitamin for all animals and humans.

Although biotin is produced in minute quantities by bacteria, plants and yeasts and is found in very small amounts in the majority of food stuffs (except egg yoke which contains 50 μg to 100 μg per egg) the recommended daily intake of around 300 μg per day is usually reached by a balanced diet and absorbed from that synthesised by intestinal microflora.

1.1.1 – Structure and Chemical Properties of Biotin

The empirical formula of biotin was determined in 1936 by Kögl ($C_{11}H_{18}O_3N_2S$) (18) and the X-ray structure elucidated 6 years later by Du Vigneaud *et al.,* which showed biotin consisted of an imidazole ring fused *cis* to a tetrahydrothiophene ring with a valeryl side chain (Figure 1.1) (19).

The structure of biotin contains three asymmetric carbons and so has eight optically active forms only one of which is biologically active, the D-(+)- biotin isomer. Current commercial synthesis is a laborious 12 step chemical process developed by Roche in 1946 (20) with significant improvements made by Lonza AG in 1988 (21). The commercial demand for the vitamin is large, so there is current industrial interest in improving the chemical synthesis as well as attempting to develop a bio-fermentation process for its production.

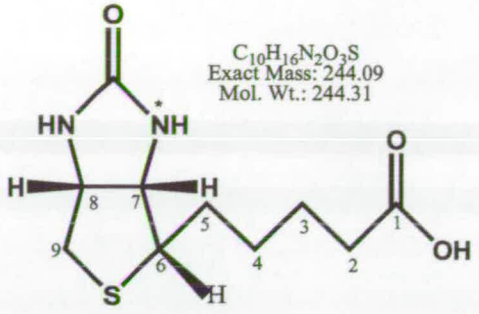


Figure 1.1 – D-(+)- biotin, consists of imidazole ring fused *cis* to a tetrahydrothiophene ring with a valeryl side chain (* = N-1 nitrogen).

Class	Enzyme	Reaction catalysed	Biochemical role
I Carboxylases	Pyruvate carboxylase (PC)	Pyruvate to oxaloacetate	Gluconeogenesis, lipogenesis
	Acetyl Co-A carboxylase (ACC)	Acetyl CoA to malonyl CoA	Fatty acid biosynthesis
	Propionyl-CoA carboxylase (PCC)	Propionyl CoA to methylmalonyl CoA	Proprionate metabolism
	3-Methylcrotonyl-CoA carboxylase (MCC)	3-Methylcrotonyl CoA to 3-methylglutaconyl CoA	Catabolism of leucine
	Geranyl-CoA carboxylase (GCC)	Geranyl CoA to carboxygeranyl CoA	Microbial catabolism of isoprenoid compounds
	Urea carboxylase (UC)	Urea to N-carboxyurea	Bacterial catabolism of urea in microbes that lack urease and grow on urea as a sole source of carbon
II. Trancarboxylases	Methylmalonyl-CoA carboxyltransferase (MMC)	Methylmalonyl CoA and pyruvate to oxaloacetate and propionyl CoA	Fermentation of certain carbohydrates to propionate in propionibacteria
III. Decarboxylases	Methylmalonyl-CoA decarboxylase	Methylmalonyl CoA to propionyl CoA and CO_2	Last step in lactate fermentation in *Micrococcus lactis*
	Oxaloacetate decarboxylase	Oxaloacetate to pyruvate and CO_2	Inducible enzyme in *Aerobacter aerogenes* challenged to grow on citrate as a carbon source

Figure 1.2 – Table of the classification of biotin utilizing enzymes, reaction catalysed and its eventual biochemical role.

1.1.2 – Biochemical roles of biotin and biotin binding proteins

The main role of biotin is as a co-factor in a number of carboxylation reactions where it functions *in vivo* as the carrier of activated CO_2 for enzymatic reactions (22).

The metabolic processes that are biotin dependent can be classified as shown in Figure 1.2. The Carboxylase or Class I enzymes form the majority of reactions involving biotin and transfer carbon dioxide derived from hydrogen carbonate to acceptor molecules/proteins. The process consists of 2 reversible steps, firstly an adenosine 5'-triphosphate (ATP) dependent step involving the formation of carboxybiotin, and subsequent transfer of this activated CO_2 to the substrate. The exact mechanism of the transfer of CO_2 to the N-1 nitrogen (Figure 1.1) of biotin to form carboxybiotin has been the subject of a number of studies (23, 24). A mechanism has been put forward which proposes that the substrates bind, and are released in a specific order, where one or more of the products are released before all the substrates are bound. The final two classes are Transcarboxylases (Class II) and Decarboxylases (Class III), in both cases there is no requirement for ATP to carboxylate biotin and active CO_2 is transferred from either an acyl derivative or a α– or β-keto ester to the biotin enzyme (25).

One of the best characterised carboxy-transfer enzymes is the acetyl-CoA carboxylase (ACC) from *E. coli*, and it has been shown to contain three subunits each of which catalyse partial reactions as described below (Figure 1.3). The *E. coli* ACC consists of three distinct subunits whose function is independent but has an absolute requirement for all components for the conversion of acetyl-CoA to malonyl-CoA using carbonate and ATP. The biotin moiety is bound to the carrier protein biotin carboxyl carrier protein (BCCP) via the biotin carboxylic acid side chain to a conserved lysine residue via an amide bond. The flexibility of this 'arm' is required to allow rotation of the biotin/carboxybiotin from the active site of subunit 2 and 3. The 2nd subunit is the biotin carboxylase (BC) which transfers CO_2 from carbonate to the N1 nitrogen of biotin at the expense of ATP. The BCCP bound carboxybiotin then delivers this activated carbon dioxide to the active site of the 3rd carboxy-transferase (CT) subunit. The CT subunit then transfers the activated CO_2 to

acetyl-CoA to form the malonyl-CoA product. The resultant protein bound biotin can then be passed back to the biotin carboxylase for further cycles (26, 27).

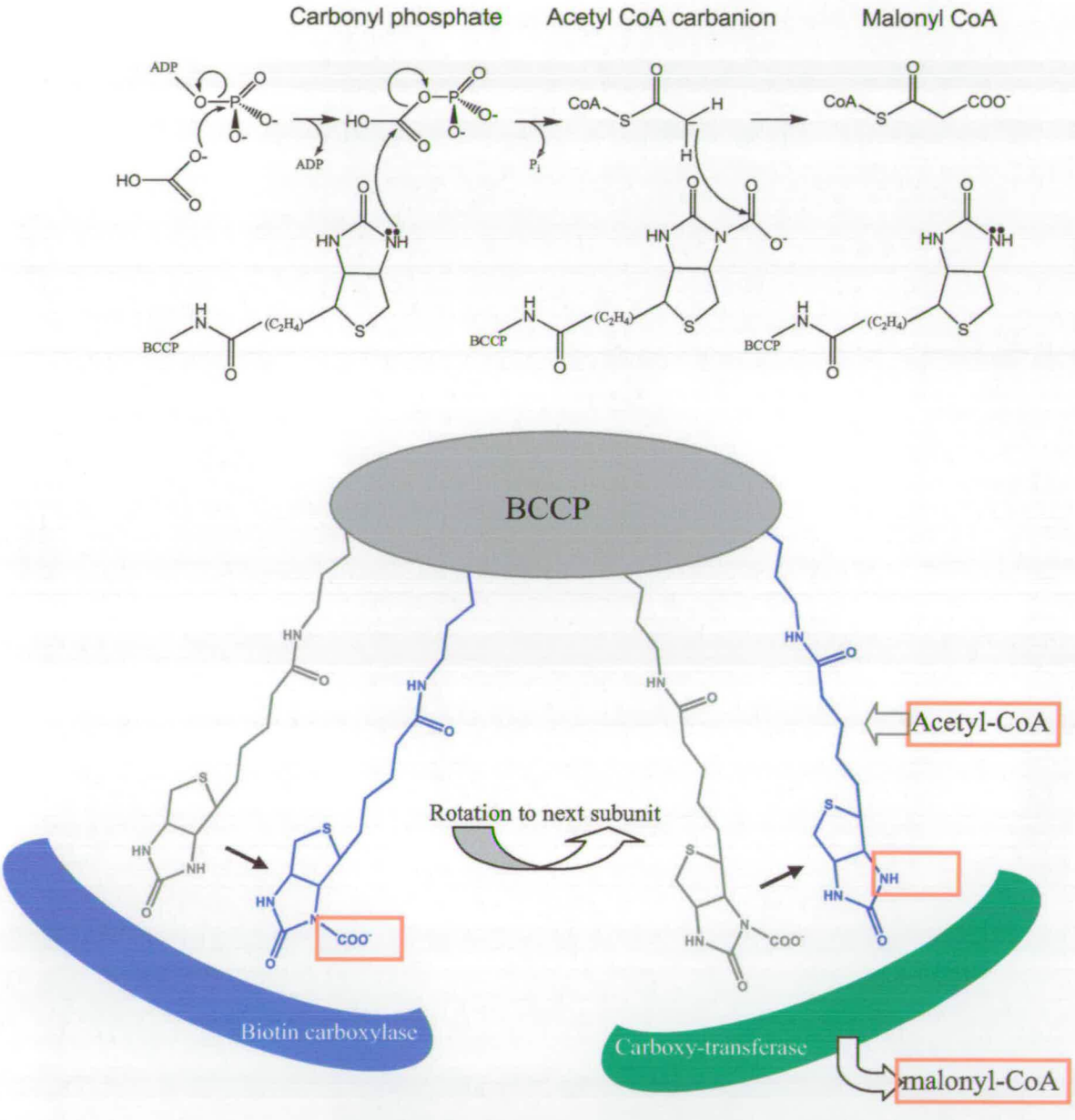


Figure 1.3 – *E. coli* acetyl-CoA carboxylase (ACC) consists of 3 protein subunits, biotin carboxyl carrier protein (BCCP), biotin carboxylase(BC) and carboxy-transferase (CT). Biotin is bound to BCCP via a conserved lysine residue which provides the flexability to translocate biotin / carboxybiotin to active sites of subunits, thus converting acetyl-CoA to malonyl-CoA.

1.1.3 – Biotin Binding Proteins

Two proteins have been shown to have very high specific binding affinities for biotin, namely the glycoprotein avidin from egg white (28) and streptavidin isolated from *Streptomyces avidinii* (29). Both proteins are homo-tetramers with monomer mass of 17.5 kDa forming a final protein mass of 70 kDa. X-ray crystallography has revealed each monomer has folded into a eight stranded anti-parallel β-barrel, where the biotin binds between two tryptophan residues at the end of the barrel (30).

Avidin to biotin affinity is one of the strongest non-covalent interactions known in biology, where the K_D of the interaction is approximately 10^{-15}M. The complex is also extremely stable to pH variations, salt concentrations and chaotrophic agents such as urea and has thus become an extremely useful affinity tag tool in molecular biology (31).

1.1.4 – Biological and Pathological Effects of Biotin Deficiency

The long term use of antibiotics can result in a reduction in intestinal microflora and this can then result in biotin deficiency (32). A lack of biotin obviously results in a lack of enzymatic activity of all biotin dependant enzyme systems as well as indirect effects such as impairment of protein and RNA synthesis. The resultant pathological effects in mammals include progressive ataxia, dermatitis, alopecia, skin lesions, nausea and metabolic acidosis. All symptoms of biotin deficiency can be reversed by supplementation of the diet with biotin.

1.2 – Biosynthesis of Biotin

1.2.1 – The *E. coli* biotin operon and regulation of biosynthesis

The majority of studies of biotin biosynthesis have been based on the *E. coli and B. sphaericus* pathways and evidence suggests that similar enzymes and intermediates are used in all microorganisms and higher plants.

The *E. coli* biotin operon (*bio* operon) is represented in Figure 1.4 and consists of a 5.8 kilobase region containing six open reading frames (ORFs) which lie between the λ attachment site and the *urv*B gene on the *E. coli* genome (33). Much of the early genetic work was carried out by del Campillo-Campbell and Rolfe and led to the identification of the genes encoded by this region (34, 35). The complementation analysis used in this work suggested that the region contains the five principle genes in biotin biosynthesis, *bioA, bioB, bioF, bioC and bioD* (36, 37). A gene, *bioH,* has also been identified that is required for biotin biosynthesis but is separately located on the genome of *E. coli* outwith the *bio* operon.

The *E. coli* biotin genes are divergently transcribed from *two pr*omoters (p*bio*BFCD and p*bio*A) (38) which are co-repressed at either side of the operator site (39, 40). The promoters are AT rich regions flanked by GC regions and it has also been established that the operator contains a hyphenated palindromic sequence from –20 to +20 that overlaps the promoters which both contain typical Pribnow box motifs of bacterial promoters.

The control of biotin biosynthesis was first recognised by Pai and Lichstein who recorded that the production of biotin decreased as extracellular biotin increased (41) and it was found that this was a repression mechanism rather than feedback inhibition (42). The regulation of gene expression within the biotin operon is achieved by the *bir*A gene product which is a 321 amino acid protein of molecular weight 35.5 kDa. It was found that BirA is bifunctional, acting as an ATP-dependent, biotin activating enzyme, converting biotin to biotinoyl-AMP (also known as biotin 5'-adenylate) and transferring this activated form of biotin to the BCCP domains of biotin dependent enzymes. BirA from *E. coli* also contains an N-terminal DNA binding domain that functions as the transcriptional regulator of the biotin operon (43).

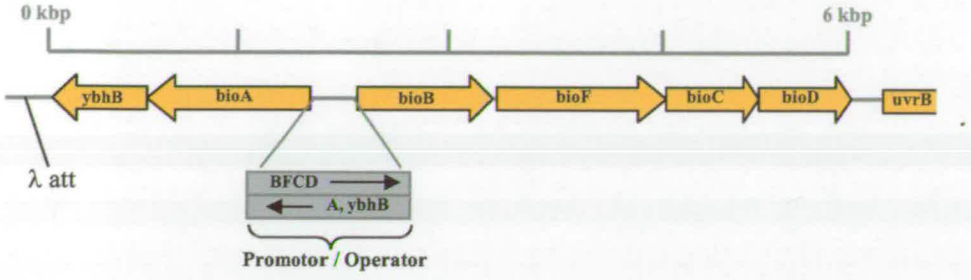


Figure 1.4 – Representation of the 6 kilobase region of the *E. coli* biotin operon, containing *bioA*,*B*,*F*,*C*, *D* & *ybhB* genes. The promoter regions lie between the *bioA* & *bioB* genes and are divergently transcribed

Figure 1.5 – Crystal structure of BirA with biotin bound (10)

The intermediate of the biotin activation reaction, BirA-biotinoyl-AMP, binds to the 40 base pair *bio* operator, repressing transcription of the genes within the operon. This is a remarkable property among DNA binding proteins as it synthesises its own co-repressor ligand. Initial structural studies by Howard suggested that BirA contains a helix-loop-helix domain similar to other DNA binding proteins based on sequence analysis (44).

This was confirmed when the X-ray crystal structure was determined in 1992 by Wilson *et al.* (10). The BirA/biotinyl-AMP complex dimerises and it is this form that is believed to be the actual DNA binding/repressor form. The structure of a BirA/biotin dimeric form was recently determined and a model for its interaction with the *bio* operator site proposed (45). The repression of the transcription of the operon thus increases proportionaly with biotin/biotinyl-AMP concentration in the growth media. This, in turn, leads to an increase in the concentration of the BirA-biotinyl-AMP repressor complex until a critical *in vivo* concentration is reached which means all operator sites are filled and transcription of the *bio* genes is fully repressed.

1.2.2 – Biosynthetic pathway in prokaryotes

The currently accepted biosynthetic pathway in prokaryotes was proposed by Eisenberg based on his studies in *E. coli* (25) and now it appears that this pathway shown in Figure 1.6 is ubiquitous in bactera.

1.2.3 – Early Steps of Biotin Biosynthesis in *E. coli*

In *B. sphaericus* it was found that the initial precursor to biotin was pimelic acid which is converted to pimeloyl-CoA by pimeloyl-CoA synthase (encoded by a bioW gene) in the presence of HSCoA, ATP and Mg^{2+} (46). However in *E. coli* no such pimeloyl-CoA synthetase activity has been detected. In contast, *E. coli* contains two genes which have been implicated in the early stages of the pathway. Results of two ^{13}C labelling studies suggest that pimeloyl-CoA is synthesised directly from malonyl-CoA by the *bio*C gene product via a pathway similar to fatty acid synthesis (47, 48). Studies by Lemone *et al.* suggest that *bio*C may catalyse the condensation

of acetate units from an initial malonyl-CoA starter unit and subsequent stepwise condensation of product and acetate units leading to formation of pimeloyl-CoA (48)

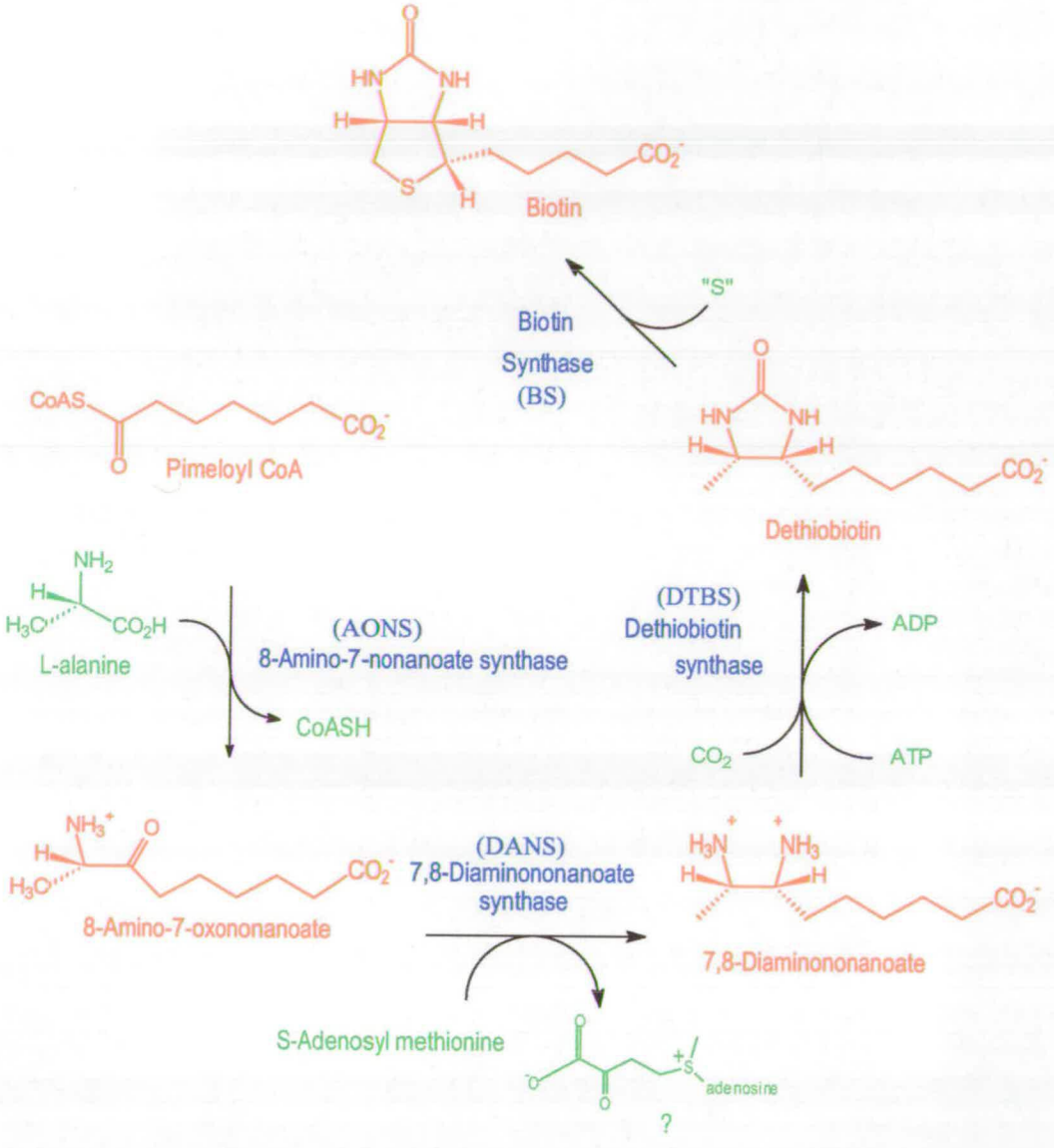


Figure 1.6 – The *E. coli* biotin biosynthetic pathway; 8-Amino-7-oxononanoate (41599Da - ANOS) is the product of the *bioF* gene, 7,8-diaminononanoate (47403Da - DANS) is the b*ioA* gene product, dethiobiotin synthase (23917kDa - DTBS) is the *bioD* gene product and biotin synthase (38648Da - BS) is the b*ioB* gene product.

The other gene implicated in this early stage is the product of the *bio*H gene and it has been suggested that this could be involved in the transfer of pimelate directily to CoA thus preventing buildup of intracellular pimelate (49). One study by Campopiano *et al.* showed that bioH binds a single CoA molecule which suggests that it may act as a scaffold for the formation of pimeloyl-CoA (50). The recent determination of the BioH x-ray crystal structure by a Canadian Structural Proteomics Consortium (shown in Figure 1.7), combined with sequence and tertiary fold analysis suggests that it could function as a carboxyesterase. Enzymatic activity was identified with a range of long-chain esters (51). This study, in conjuction with our findings, suggest that BioH plays an enzymatic role in the formation of pimeloyl-CoA but the precise details of the substrates involved within this putative system are yet to be elucidated. A search for protein partners of BioH using antibody pull down experiments may reveal some important incites to the exact cellular role.

1.2.4 – *E. coli* 8-Amino-7-oxononanoate synthase (AONS)

After formation of the key precursor pimeloyl-CoA, biotin biosynthesis in all organisms proceeds in a similar manner. Pimeloyl-CoA is converted to the next intermediate in the pathway, 8-Amino-7-oxononanoate (AON) by 8-Amino-7-oxononanoate synthase (AONS), a 384 amino acid protein with molecular weight of 41.5 kDa. This enzyme catalyses the decarboxylative condensation of L-alanine with pimeloyl-CoA to form AON, CO_2 and HSCoA. The catalytically active homo-dimeric form has an absolute requirement for a pyridoxal 5'-phosphate (PLP) cofactor for enzymatic activity. The *bioF* genes from *E. coli* and *B. sphaericus* have both been expressed in *E. coli*, protein purified, (52, 12) and the structure of the *E. coli* protein has been determined by the Edinburgh group (PDB 1BSO) (12) as shown in Figure 1.8. The protein is a homodimer with an overall structure typical of a type II aminotransferase (53). The AONS monomer has 3 domains; the N-terminus wraps around the opposing monomer, the second is a large central catalytic domain and a third small C-terminal domain. The dimer contains two active sites with residues contributed from each monomer and it is located in a cleft between the 2nd and 3rd domains. The PLP cofactor is bound via lysine residue (K236) which is conserved

among all AONS enzymes studied to date. The structure of the product AON-bound form revealed the residues involved in catalysis (Webster *et al.* (54)).

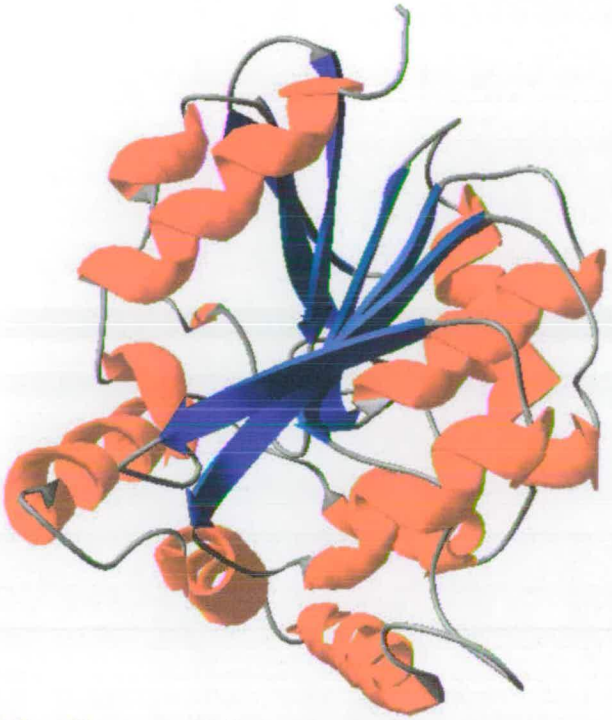


Figure 1.7 – X-ray crystal structure of the monomeric *E. coli* BioH (PDB accession code: 1M33). The catalytic Rossman fold consists of seven strand β sheets (blue) surrounded with α helices (red)

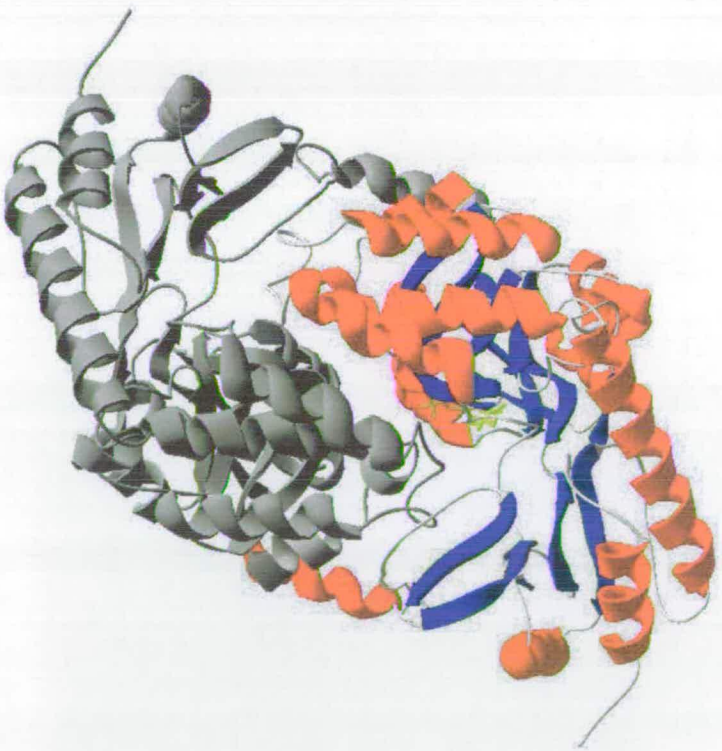


Figure 1.8 – The x-ray crystal structure of internal aldimine of *E. coli* AONS (1BSO). The dimer is shown, with a single chain coloured. Monomer contains 3 domains the N-terminus wraps around the opposing monomer. The PLP co-factor is bound as an aldimine to lysine 236 shown in yellow (12).

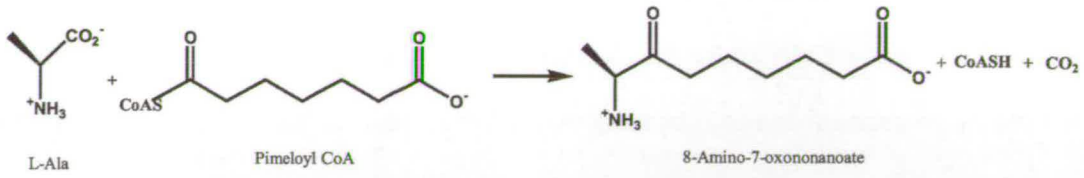


Figure 1.9 – Reaction catalysed by 8-Amino-7-oxoxnonanoate synthase. The carboxylate group of alanine is replaced by a pimeloyl moiety.

AONS is a member of a family of enzymes which catalyse similar chemistry – the α-oxoamine synthase family. Other members include SPT the first enzyme in the sphingolipid biosynthetic pathway, ALAS involved in heme biosynthesis and KBL which is involved in the metabolism of threonine.

1.2.5 – *E. coli* 7,8-diaminononanoate synthase (DANS)

After the formation of AON by AONS the next step in the biosynthetic pathway is a transamination reaction which is catalysed by 7,8-Diaminononanoate synthase (DANS) which leads to the formation of 7,8-diaminononanoate (DAN). This enzyme is the product of the *bio*A gene which encodes a 430 amino acid protein with monomer molecular weight of 47.4 kDa. The dimeric enzyme has a requirement for PLP, and unusually, the amino donor is S-adenosyl methionine (SAM). The DANS enzyme is unique in its requirement for SAM but resembles both AONS and other type II aminotransferase structures in that the homodimer is formed from monomers with three domains very similar to those described for AONS. The X-ray crystal structure has been determined by Lindquist (55) and Alexeev (2) and that solved by Alexeev is shown in Figure 1.10. Recent mutational studies have explored the residues involved in catalysis. Four residues were found to be important for the complete reaction, two tyrosine residues were suggested to be involved exclusively with the DAPA half reaction, an arginine residue which is involved in the SAM half reaction and an asparagine residue, the mutation of which inhibited both half reactions (56). However, the exact details of SAM binding are still unclear since a structure of the DANS-SAM external aldimine complex has remained elusive.

Interestingly, recent work by Marquet and coworkers has revealed the structural details of the inhibition of DANS by the antibiotic amiclenomycin.

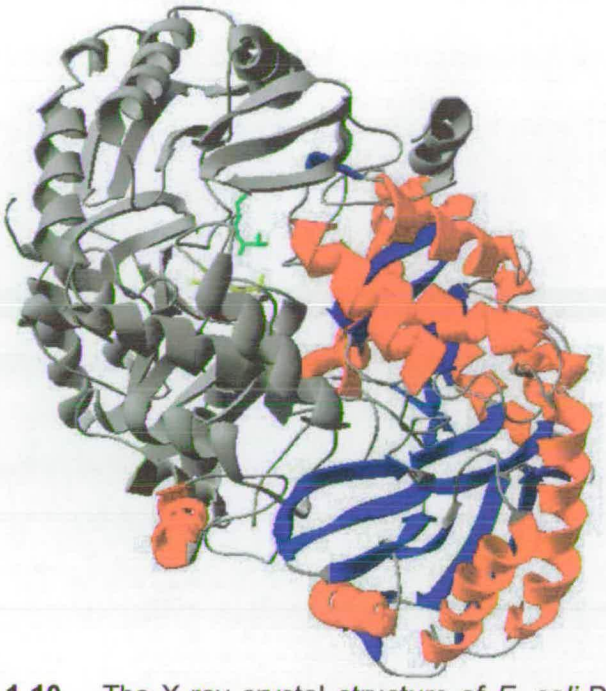


Figure 1.10 – The X-ray crystal structure of *E. coli* DANS dimer. Each monomer is coloured with β sheets blue and α helices red and contains 3 domains the first of which wraps around the opposing monomer. The PLP co-factor is bound to lysine 274 (2).

Figure 1.11 – The X-Ray crystal structure of *E. coli* DTBS (1DAD). Dimer shown, with one monomer coloured β sheets blue and α helices red. (15)

The inhibition of DANS with this antibiotic was found to be specific for the *cis* isomer and it is suggested by the authors that PLP forms a covalent bond with the inhibitor and is initiated by hydrogen abstraction from the cyclohexadiene ring of amiclenomycin by the sidechain of the lysine residue (K274) lying in close proximity (56).

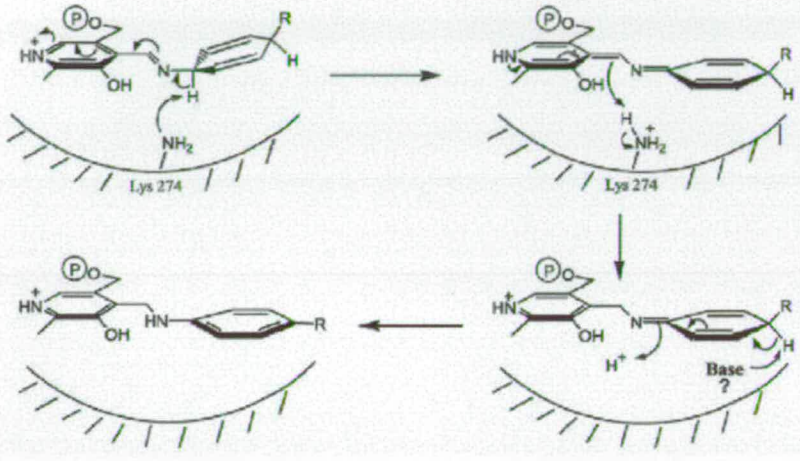


Figure 1.11b – Proposed mechanism of formation of the covalent aromatic adduct between amiclenomycin and PLP in the active site of DAPA synthase.

1.2.6 – *E. coli* Dethiobiotin Synthase (DTBS)

The penultimate step in the pathway is catalyzed by the dethiobiotin synthetase (DTBS) enzyme encoded by the *bio*D gene. The product of the bioD gene is a 224 amino acid protein, where the N-term methionine is cleaved after translation with monomer molecular mass of 23.8 kDa. The catalytically active, dimeric form of DTBS catalyses the insertion of carbon monoxide into DAN forming dethiobiotin and the reaction has a requirment for ATP and Mg^{2+} (57). Interestingly this enzyme is unusual in that it does not use biotin as a CO_2 carrier but instead uses ATP to drive the formation of the two C-N bonds of the ureido ring. The *E. coli* DTBS shows little sequence homology with other proteins except related DTBS proteins from *Serratia marcescens* and *B. sphaericus* (58, 59) with which it shares 70.2% and 49% sequence identity respectivly. The ligand-free crystal structure was determined to a

resolution 1.6Å (1DAD) (15, 60) as shown in Figure 1.11. Each monomer of the dimer has seven stranded parallel β sheets surrounded by helices. These sheets contain the classic mononucleotide binding motif G-X_5-G-K-T and the sequence of catalytic steps have been examined in depth and roles of these residues elucidated (61, 62, 63). The DuPont group have subsequently determined the structures of DTBS with a range of intermediates, substrate analogues and inhibitors which have provided insight into the catalytic cycle (64).

1.2.7 – The Final Step: Dethiobiotin – Biotin conversion

The final step in the biosynthetic pathway is chemically the most challenging and has been the subject of intensive research by a number of groups over the past decade. This step involves the insertion of a sulfur atom between the C-6 and C-9 carbons of DTB, thus forming two carbon-sulfur bonds of the tetrahydrothiophene ring of biotin as shown in Figure 1.12. In absolute enzymological terms this step is the least well understood regardless of the large amount of work that has been undertaken. It is accepted that the *bioB* gene product, biotin synthase, is responsible for this conversion, however to date, enzymatic turnover of biotin synthase has not been achieved *in vitro*. Similarity searches have revealed a number of sequence homologs in bacteria and plants to that of the predicted *E. coli* biotin synthase. The entire sequence from twenty six species are known of which only four species, *B. sphaericus* (65), *A. thaliana* (66), *S. cerevisiae* (67) and *E. coli* (68) have been cloned, overexpressed and purified. Other work has been sufficient to verify that genes of the species *Methylobacillus*, *B. flavum* (69), *B. subilis* (70), *E. herbicola* (71) and *S. marcescens* (58) encode viable active biotin synthase by complementation analysis.

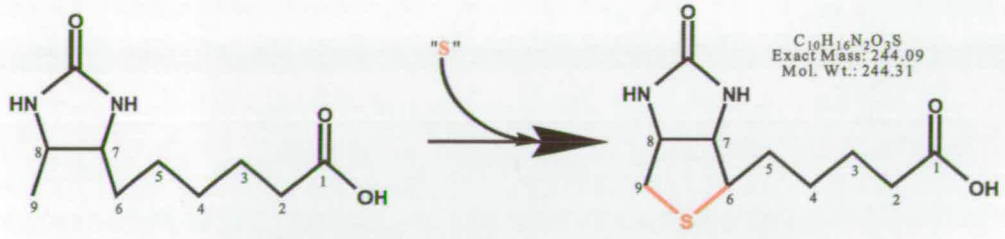


Figure 1.12 – Conversion of dethiobiotin to biotin thought to be catalysed by biotin synthase forms two new carbon-sulfur bonds (red).

1.3 – *E. coli bioB* gene product: Biotin synthase (BS)

The *E. coli bioB* gene encodes a biotin synthase (BS) polypeptide of 346 amino acids with a predicted molecular weight of 38.6 kDa (72). The *E. coli* gene was initially cloned and overexpressed by Sanyal *et al.* and they were the first group to achieve the purification of native, un-tagged BS. They identified two species by native gel electrophoresis with masses of approximately 82kDa and 104kDa as judged by SDS-PAGE which were shown to be homodimers. Iron and sulfur analysis revealed the BS:sulfur:iron stoichiometry in each species and suggested that they contained two [2Fe-2S] clusters and one [2Fe-2S] cluster respectively. The BS with apparent mass of approximately 104kDa was converted to the 82kDa species upon incubation with Fe^{3+}, S^{2-} and β-mercaptoethanol. Both species were red/brown in colour and had UV-Vis maxima at 324nm and 419nm which is consistent with other [Fe-S] cluster proteins (68). One further species of biotin synthase has been reported with apparent molecular weight of 160kDa which again showed similar UV-Vis maxima as before (67).

1.3.1 Biotin synthase – a member of the Radical SAM superfamily

Multiple sequence alignment of biotin synthase homologs reveals several regions of high identity and others of high conservation. The region between lysine49 and serine63 shows very high conservation with three cysteine residues being completely conserved within the motif -**C**-X_3-**C**-X_2-**C**- which was originally known as the Cys box motif (67). In a seminal bioinformatic study in 2001, Sofia and colleagues discovered a protein superfamily with >600 members which contain this motif – named the "Radical SAM superfamily" (72). The proteins in this family bind SAM and generate a reactive species, the 5'deoxyadenosine radical, which is then used in 1 electron radical chemistry in a number of diverse biochemical pathways (73). For example, the signature motif has been identified in anaerobic ribonucleotide reductase activating enzyme (ARR-AE, Swissprot acc number P39329) (74), benzylsuccinate synthase activating enzyme (CAA05050) (75), lipoic acid synthase (P25845) (76), lysine aminomutase LAM (77), the nitrogen fixation B gene product (P11067) (78), FNR protein (P03019) (79),

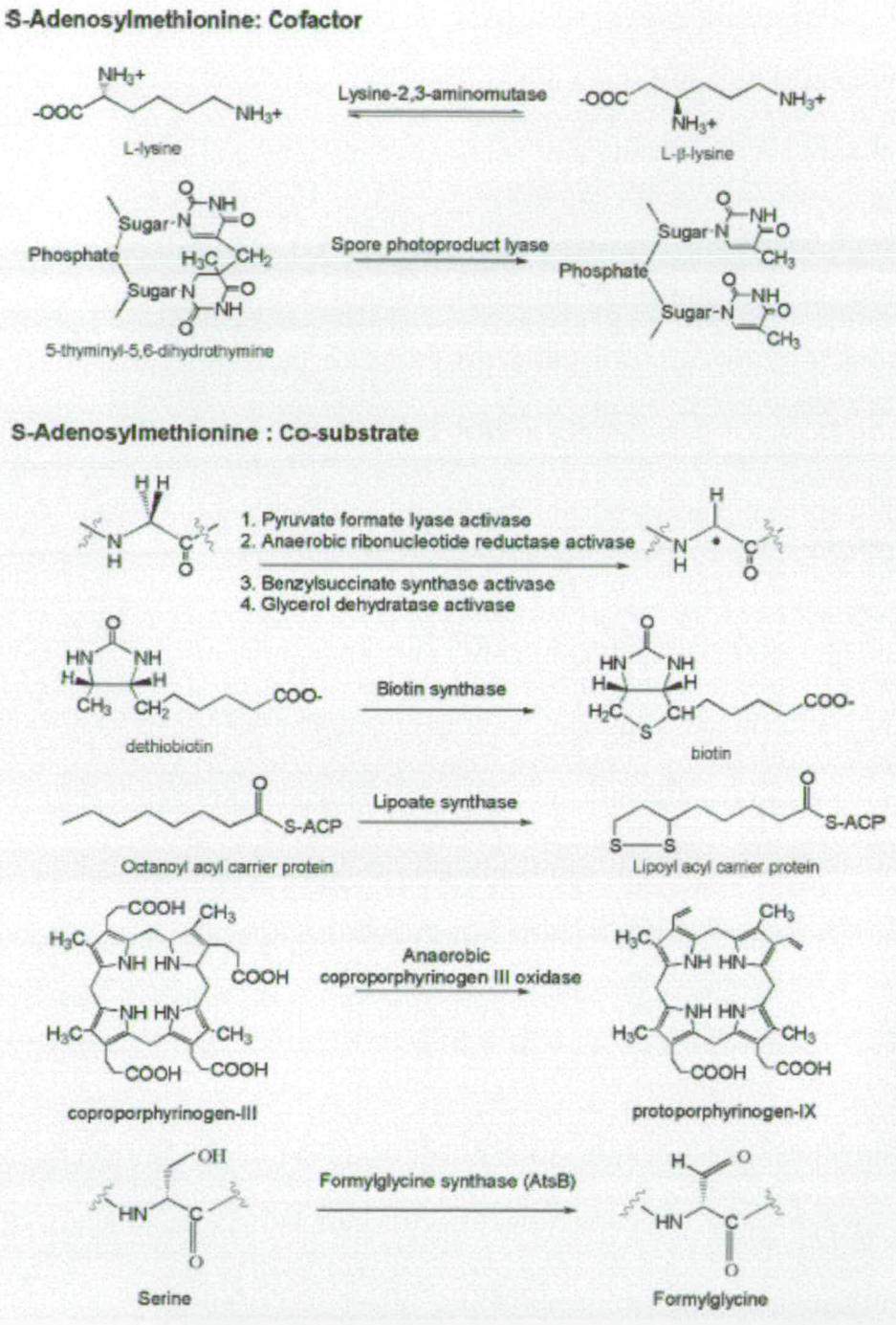


Figure 1.13 – Some reactions catalysed by some members of the SAM radical super family. Reproduced from "S-Adenosylmethionine radical enzymes" Marja S. Huhta (5).

pyrrolo-quinoline-quinone synthase (Q01060) (80), pyruvate formate lyase activating protein (PFL-AE) (P09374) (81), spore photoproduct lyase (P37956) (82), molybdenum cofactor synthase (MOAA) (O27593) (83), NARA protein (P39757) (84) and the *thiH* gene product (P30140) (85). It was recognised that all these proteins bound [Fe-S] clusters which led to speculation that the conserved cysteine residues within the Cys box were the ligands for cluster binding. Evidence for this was shown by mutagenesis studies that indeed the Cys53, Cys57 and Cys60 of the *E. coli* biotin synthase protein are involved in the coordination of an Fe-S cluster (86, 87).

1.3.2 – Biotin synthase – controversies of cofactor composition

Since the initial purification of BS in 1994 seven groups have carried out extensive studies on *E. coli* BS. Methods for the expression, purification, handling and characterisation vary between laboratories. Each laboratory produces BS with varying activity as well as differing Fe, S and cofactor content. No single method has been adopted within the field but there is general agreement that active BS contains two types of Fe-S cluster bound at different sites on the enzyme using conserved cysteine residues. The Fe-S clusters in biotin synthase are also labile which has also led to preparations varying in Fe and S content. While aerobic purification of the protein produces a dimer with two $[2Fe\text{-}2S]^{2+}$ clusters (8, 88) the protein can be reconstituted anaerobically to contain mainly $[4Fe\text{-}4S]^{2+}$ or $[4Fe\text{-}4S]^{+}$ clusters upon strong chemical reduction with dithionite with or without, respectively, 60% ethylene glycol (8, 88, 89, 90). Also reduction under milder conditions appeared to produce a 1:1 ratio of $[2Fe\text{-}2S]^{2+}$:$[4Fe\text{-}4S]^{2+}$ clusters as judged by UV-Vis and chemical analysis (7) and comparisons of amounts of DTB to biotin conversion suggested that this last state provides highest activity however it is still unclear as to the processes involved in these conversions. A proposed mechanism for interconversions of 2 x [2Fe-2S] clusters to a single [4Fe-4S] cluster was proposed by Johnson *et al.* where a single [2Fe-2S] from each monomer of the biotin synthase dimer contributes to the final [4Fe-4S] cluster and is shown in Figure 1.15 below (8)

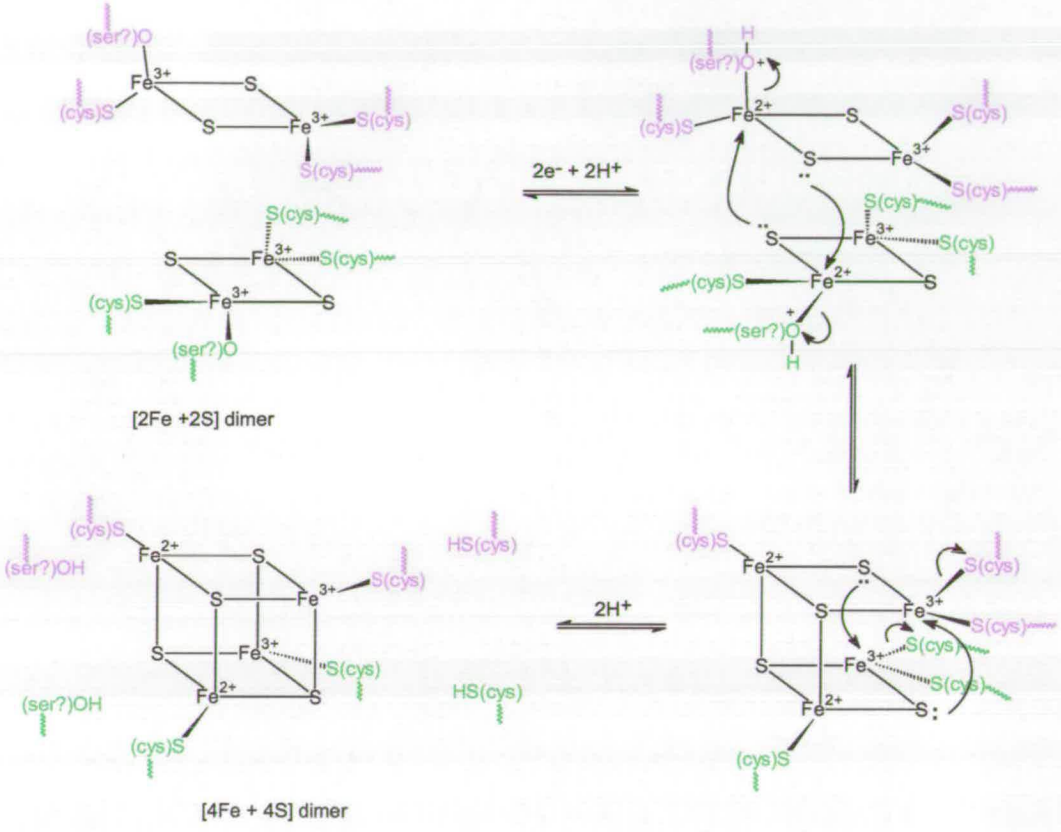


Figure 1.15 – Scheme proposed by Johnson *et al.* for conversion of 2[2Fe-2S] clusters, one from each monomer, to a single [4Fe-4S] cluster. (8)

The field was further divided by the finding by Fontecave's laboratory that BS displays PLP-dependent cysteine desulfurase activity and a BS-bound persulfide was implicated in the sulfur transfer step (91, 92, 93). This paper has garnered little or no support from other groups in the field and as such is the subject of controversy (see below). This thesis describes efforts to clarify this area further.

It was presumed that BS preparations containing $[2Fe\text{-}2S]^{2+}$ clusters are inactive forms of the protein produced as an artefact of overexpression under aerobic growth conditions and purification procedures. Recent groundbreaking *in vivo* Mössbauer studies which measured the spectra of whole cells expressing BS found that the main product of *bioB* overexpression was in fact the $[2Fe\text{-}2S]^{2+}$ cluster. The appearance of this species within the cell before isolation led Johnson and colleagues to conclude that this BS form was the product of hyperexpression of the gene in *E. coli* in the absence of Fe-S cluster enzymes which leads to an inactive form *in vivo* (94). Although extensive work has been carried out to ascertain the nature of the cluster(s) present in active biotin synthase it still remains a highly contentious issue and exact mechanistic roles for the cluster(s) have yet to be agreed between experts in this field.

Although much has been discovered about biotin synthase in the last 20 years certain issues still remain unclear such as the exact source of sulfur, the mechanism of formation of each of the two C-S bonds and an explanation as to why there have been no preparations of BS capable of multiple turnovers *in vitro*. It is now accepted that biotin synthase is a SAM dependant enzyme that also requires NADPH, flavodoxin (FLAV) and flavodoxin reductase (FRED) to supply the electrons for cluster reduction, subsequent SAM cleavage and biotin formation (95, 96, 97, 98). This system provides maximum activity *in vitro* however it has also been shown that photoreduced deazaflavin can also be used (99). Whether the FLAV/FRED reducing system is indeed the true *in vivo* redox partner is also open to conjecture since addition of these enzymes still does not produce the desired *in vitro* catalytic DTB to biotin activity.

1.3.3 – Radical SAM-dependent mechanism of BS

The binding of SAM to BS has been investigated by Mössbauer and Raman spectroscopy by examining spectral changes observed upon addition of SAM to BS (94). This led to the development of some insightful models which suggested that SAM was bound directly to the [4Fe-4S] cluster, later this was confirmed by X-Ray crystallography (see Section 1.3.4). It is suggested that one electron reduction occurs after SAM binding resulting formation of a deoxyadenosyl radical. This DOA radical is then proposed to abstract the hydrogen from C6 and/or C9 and the resultant carbon radical is then used to form the first of the C-S bonds. Since a [4Fe-4S] cluster requires four thiol ligands it was thought that three of the ligands were provided by the conserved cysteine residues in the 'cys-box' $-C-X_3-C-X_2-C-$ motif and that the fourth iron ligand in the cluster is provided by an atom from SAM. It also follows that for every SAM molecular turned over an equivalent amount of methionine and 5'-deoxyadenosine are produced. In two separate studies a 1:1 ratio of methionine: 5'-deoxyadenosine were observed but this was 3 times as much with respect to the amount of biotin produced (100, 101, 102). This discrepancy was put down to abortive SAM cleavage, without biotin production. Both studies concluded that these results were consistent with consumption of 2 SAM molecules per DTB to biotin conversion and therefore formation of one 5'-deoxy adenosine radical per C-S bond. Calculations of SAM consumption are difficult and prone to error due to the assays being performed using crude cell extracts or having extremely low turnover activities compared with previously reported values (100, 101, 102).

In contrast to these earlier studies, it has recently been reported that only one SAM is required for biotin formation when an assay was performed using a biotin synthase preparation containing PLP and a single [4Fe-4S] cluster (91, 92). This work showed that an *in vitro* assay system containing purified BS, PLP, DTB, DTT, SAM, FLAV, FRED, NADPH and cysteine had biotin production activity equivalent to the reported maximum of one biotin molecule per molecule of BS. This single turnover was achieved when PLP was included in the assay system compared to 0.25 biotin molecule per molecule of BS without PLP (92). Fontecave and colleagues also measured a biotin to methionine ratio of 1:1 suggesting that only one SAM radical cleavage event is required for the production of 1 biotin molecule (91). The

conclusions drawn from this study were that biotin synthase is capable of liberating sulfur from cysteine by a NifS-like, PLP-dependent cysteine desulfurase activity. The conversion of cysteine to alanine was confirmed by isolation of the alanine as a ninhydrin derivative and comparison with amino acid standards on HPLC. This cysteine desulfurase activity of biotin synthase was also shown to exhibit multiple turnovers albeit at slow rates in the region of 1 per hour. This system also supported a second DTB to biotin conversion if the proteins were rigorously washed of any low molecular weight products. The reasoning behind this wash step was that the 5'-deoxyadenosine (DOAH) produced after a single turnover is a potent inhibitor of the BS reaction and its removal is essential for multiple turnovers (91). Again, the role of PLP and the inhibition by DOAH is the subject of argument by others in the field.

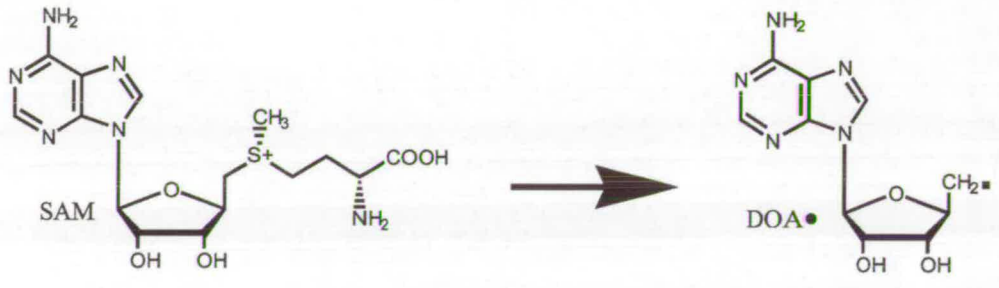


Figure 1.16 – Reductive cleavage of S-adenosylmethionine to the 5'-deoxyadenosine radical facilitated by the reduced [4Fe-4S] cluster of the radical SAM superfamily (5).

1.3.4 – Source of sulfur incorporated into biotin

Cysteine is accepted as the source of the sulfur atom in biotin but the details of the pathway involved in the breaking of the cysteine C-S bond and the subsequent intermediates involved during its transfer to DTB have been the subject of numerous studies. This has led two theories of sulfur incorporation; the first proposes that the cysteine sulfur is incorporated into the [2Fe-2S] cluster and it is a sulfur from this species that is the direct source of the sulfur in biotin (7, 103). The second theory is that the cysteine is incorporated in a PLP-dependent mechanism into a cysteine-bound persulfide (S-S) (93, 91, 92).

The background which led to these two hypotheses is as follows; crude bacterial extracts containing ^{35}S cysteine were shown to yield biotin with ^{35}S label but with significantly lower levels than expected indicating other sources of sulfur are used in this system (102, 65). Several others have reported that in a more controlled system containing purified biotin synthase no transfer of the ^{35}S label into biotin was observed unless a cysteine desulfurase enzyme (NifS, IscS) is included in the system (95, 104, 105). As well as this it has been demonstrated that ^{34}S or ^{35}S has been transferred from labelled Na_2S suggesting that S^{2-} can be used as the source of sulfide *in vitro*. In an attempt to show the source of 'S' inserted into biotin originates from the protein bound clusters, preparations of apo-biotin synthase were reconstituted with $Fe^{2+}/^{34}S^{2-}$ or with $Fe^{2+}/(^{35}S)$-cysteine and cysteine desulfurase (NifS) however incorporation of the labelled S was found to be only 62% and 56% respectively (106, 104) suggesting that large portions of ^{32}S remain within the protein as sulfane type species which are clearly distinct from the bound cluster(s).

The hypothesis initially proposed by Flint and colleagues and then enhanced by the labelling experiments from the Marquet laboratory states that the sulfur atom inserted into biotin originates from the cluster itself and is derived from one of the cluster bridging sulfur atoms (106, 68). However other sulfur sources could not be excluded as treatment of BS with labelled sulfur could affect other sulfur containing species within the protein. This theory was later revised by Jarrett and co-workers who suggested that a second cluster within each monomer of biotin synthase was required (7, 107, 108).

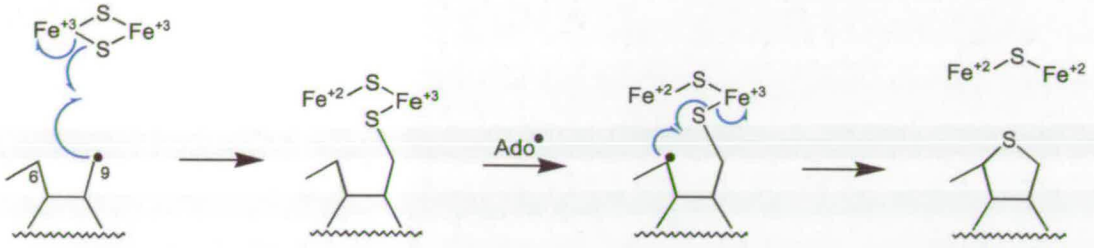


Figure 1.17 – Hypothetical mechanism of sulfur insertion to form biotin where two SAM molecules are consumed and the second [2Fe-2S] cluster is sulfur donor (7).

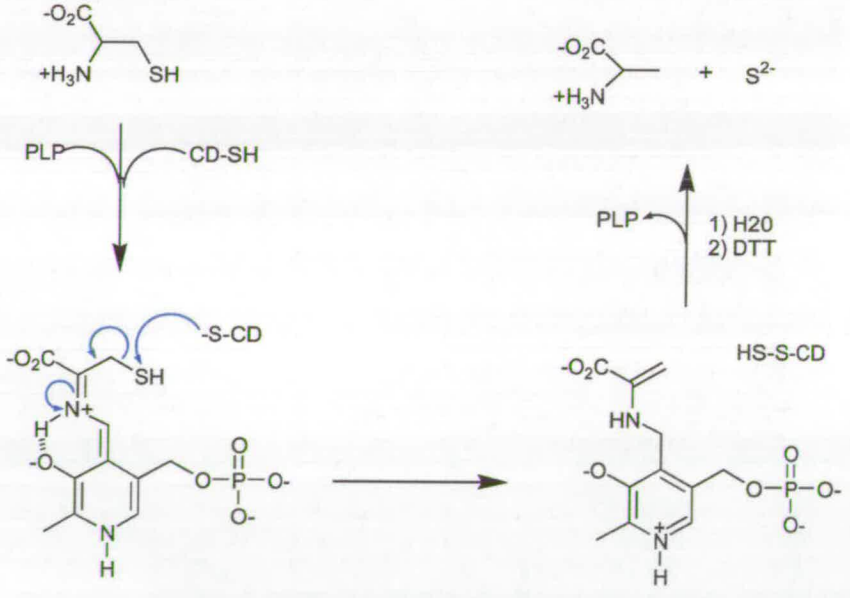


Figure 1.18 – Proposed mechanism of PLP dependant cysteine desulfurase activity of biotin synthase. Reproduced from Fontecave *et al.* (9).

This hypothesis was based on the already recognised reduced $[4Fe\text{-}4S]^{1+}$ cluster acting as the electron source to cleave SAM and also a second [2Fe-2S] cluster which is responsible for donation of the sulfur species to form biotin via the mechanism proposed in Figure 1.17. The mechanism suggests initial abstraction of a hydrogen atom from C9 to form a radical which couples to sulfur in the cluster. Ring closure occurs after generation of a second radical intermediate formed from another SAM radical cleavage thus resulting in formation of biotin. This mechanism suggests that the [2Fe-2S] cluster is destroyed during the formation of biotin and spectroscopic evidence for this was obtained by Jarrett and colleagues.

In contrast, Fontecave and colleagues have proposed a [2Fe-2S] cluster-independent mechanism. They have suggested that biotin synthase can liberate S^{2-} by a similar mechanism to other PLP-dependent cysteine desulfurase enzymes (fig. 1.18). This multi-step mechanism involves the formation of a substrate cysteine-PLP quinonoid adduct and subsequent nucelophilic attack by a thiolate anion of protein bound cysteine (91). The exact PLP-binding residue is postulated to be lysine 49 but the exact role of this residue is the subject of this present study. This generates a PLP bound enamine which is subsequently hydrolysed to alanine and a S^{2-} species (fig. 1.18).

The resultant S^{2-} is then proposed to form a protein bound persulfide which is the direct sulfur donor to the activated DTB radical produced by the [4Fe-4S]/SAM radical (fig. 1.19).

The PLP-dependent mechanism involves the formation of unprecedented persulfo-metal complex as no examples have been observed in other metallo-proteins. The mechanism consists of the coupling of the C9-radical and the terminal sulfur of the persulfide giving rise to a thiyl radical and the iron bound dehydrogenated 9-mercaptodethiobiotin. The thiyl radical is postulated to abstract the hydrogen at C6 and ring closure occurs with a one-electron reduction of the cluster (9).

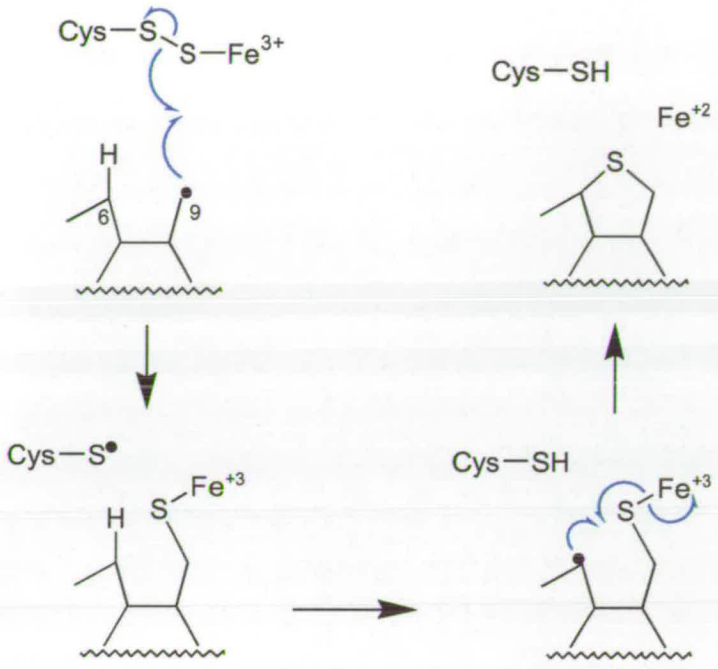


Figure 1.19 – Proposed mechanism of protein bound persulfide as donor to convert activated DTB radical to biotin (9).

This biotin production mechanism involving the PLP-dependant cysteine desulfurase activity is still highly debated in the literature as a direct alternative to the [2Fe-2S] cluster mechanism discussed previously. However it is interesting to note several similarities between them. Firstly, both have suggested that the final acceptor of the electron that results from the formation of biotin is ferric iron. Secondly, the authors of both mechanisms implicate the same conserved cysteine residues within the biotin synthase sequence as being essential for activity but provide differing explanations. Jarrett and colleagues proposed that the conserved cysteine residues, Cys97, Cys128 and Cys188 which have been shown to be catalytically important (109) are involved in the binding of the second cluster, namely the [2Fe-2S] sulfur donating cluster. However, the same conserved residues (Cys97, Cys128 and Cys188) have been shown to be extremely important (C188 less so) for the PLP-dependent cysteine desulfurase activity (92).

Given the complexity of the biotin synthase system and despite the determination of the crystal structure of reconstituted BS (see Section 1.3.5) it is still difficult to discount the published work in favour of one mechanism or the other. The fact that two different preparations of biotin synthase, one with [4Fe-4S] and [2Fe-2S] clusters and the other with a single [4Fe-4S] cluster and PLP bound, seem to produce very similar biotin production rates, suggests that a number of possibly similar, but hitherto uncharacterised, intermediates have still to be revealed in the sulfur insertion reaction.

1.3.5 – X-ray crystal structure of *E. coli* biotin synthase

Until recently no 3D structures had been determined of any member of the radical-SAM superfamily of enzymes presumably because it had been tremendously difficult to obtain a solution of the protein in a suitably homogeneous state to enable crystallisation. However, a milestone was reached in this field by the publication of the x-ray crystal structure of a reconstituted *E. coli* BS at 3.4Å resolution through the collaborative efforts of the Jarrett and Drennan groups (16). The structure of BS was determined using a single crystal of a reconstituted protein preparation via the hanging drop method in an Argon atmosphere. Biotin synthase is a homo-dimer (Figure 1.20) and the BS monomer tertiary fold was not found to be unique but is in fact a common triosephosphate isomerase (TIM) type $(\alpha/\beta)_8$ barrel with two additional α-helices at the N-terminus and a disordered region at the C-terminus (16). It is interesting to note that BS shares the same fold as three AdoCbl-dependant enzymes that also utilise radical-chemistry mechanisms (110) but it is unusual to find [Fe-S] clusters bound in this type of fold. As shown in Figure 1.21b the authors identified two clusters bound ~16Å apart (a single [4Fe-4S] and a single [2Fe-2S]). The [4Fe-4S] cluster is found toward the C-terminal end of the TIM barrel at least 30Å from the dimer interface whereas the [2Fe-2S] cluster is located deep within the barrel 25Å from the C-terminus. The authors were also delighted to find not only SAM bound directly to the [4Fe-4S] cluster but also DTB within a few angstroms of the cluster, poised for hydrogen abstraction. They found no evidence of PLP bound to BS, however no PLP was

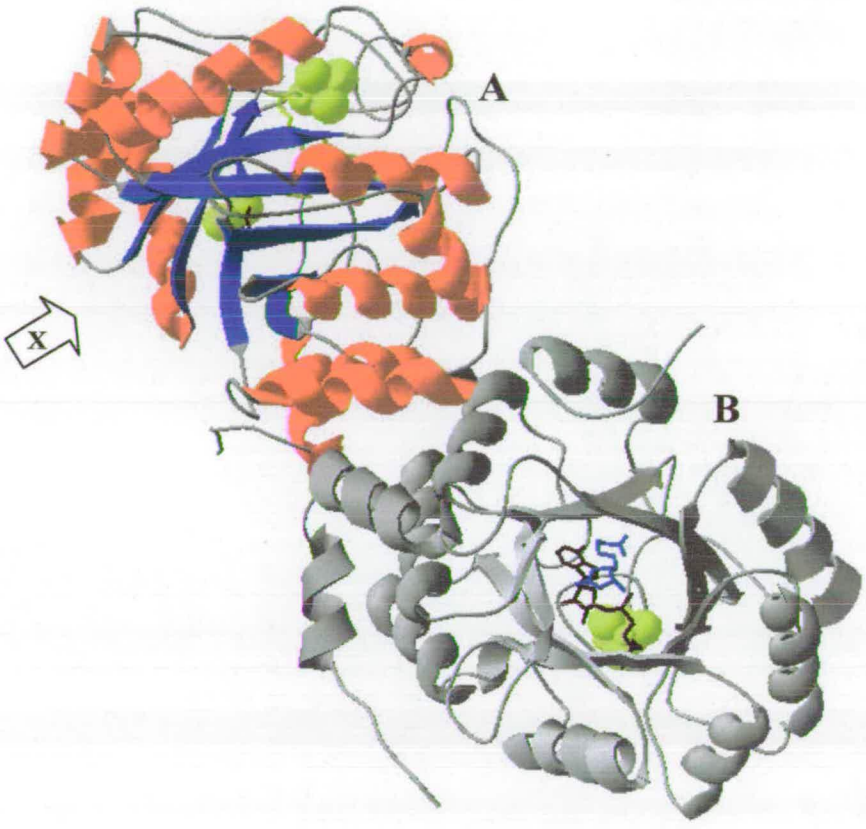


Figure 1.20 – X-Ray crystal dimeric structure of *E. coli* biotin synthase (1R30) as solved by Drennan and co workers. Monomer A is shown with α-helicies in **red** and β-sheets in **blue**. The relative positions of [4Fe-4S] and [2Fe-2S] clusters (yellow spacefill) are shown with relatively exposed [4Fe-4S] cluster and buried [2Fe-2S] cluster. Monomer B is viewed as indicated by arrow (**X**) [4Fe-4S] in spacefill and yellow with [2Fe-2S] cluster omitted for clarity. Also shown is SAM (**black**) and DTB (**blue**) .(16)

added during the reconstitution process in contrast to the 5mM concentration of SAM and DTB in the reconstitution buffer (111). Although the authors do mention the only conserved lysine residue in biotin synthase (K49) as a putative site of PLP modification they disregard this theory due to its solvent exposed position. The discovery that BS has a very common 3D topology, rather than a unique tertiary fold was slightly disappointing. However, it is not surprising since such a common scaffold is used in a diverse range of enzymatic reactions. The crystal structure has finally revealed details of important residues involved in the mechanism, the binding sites of the Fe-S clusters, SAM and DTB substrates and their spatial positions relative to each other. Figure 1.21b shows both the [4Fe-4S] and the [2Fe-2S] clusters and their corresponding interacting residues. The major binding interactions within the active site are summarised in Figure 1.2a. The [4Fe-4S] cluster is bound by Cys53, 57 and 60 as previously postulated by mutational studies (87, 86) and the final iron is bound via a N/O bidentate contact with the amino-group nitrogen and carboxyl-group oxygen of SAM predicted by spectroscopic studies (16). The remainder of the SAM molecule interacts with Arg173, Asn153, Asp155, Ile192 and Tyr59 as shown. The second cluster, [2Fe-2S], is bound by Cys97, 128 and 188 which have all been implicated as vital for biotin synthase activity (109) and for the putative cysteine desulfurase activity (91). A final unusual arginine ligand (Arg260) is shown to interact with the [2Fe-2S] cluster and this Arg260 also is in contact with residues Ser43, Ser218, Ser283 and Arg95 as shown. The DTB substrate is shown to interact with the conserved residues Asn222, Asn151 and Asn153 in the active site and also the carboxylate tail of DTB stacking with the adenine ring of SAM. The carboxylate tail also appears to interact with residues Thr292 and Thr293.

The binding of substrates SAM and DTB is of a cooperative fashion both displaying significant interactions with each other as well as the protein chain. SAM is suitably positioned in this structure to accept the electron from the [4Fe-4S] cluster directly through the N/O coordination of the bridging Fe. The close proximity of the cluster to the surface of the protein, within 7Å, suggests that a system like NADPH, FLAV and FRED could directly supply an electron for this one electron reduction. The formation of the $DOA^{\bullet}$ radical within this active site would result in the radical being 3.9Å and 4.1Å from the C9 and C6 positions respectively of DTB and so it is

possible that abstraction of hydrogen from either position could be achieved by the DOA$^{\bullet}$ radical.

The final controversial step of the mechanism is the insertion of S, and with respect to the second cluster the bridging Fe is within 4.6Å of C9 of DTB and as such could act as donor. If the second cluster is indeed found to be the donor two possibilities remain. Firstly biotin synthase acts as a suicide enzyme as suggested by Drennan and co-workers (16) or the cluster can be rebuilt either by a recognised system such as ISC or perhaps by biotin synthase itself in which case this would implicate the PLP dependant cysteine desulfurase activity reported (92, 91) as vital to this rebuilding process. Although this cluster is buried within the barrel movement of Arg95 and/or Tyr149 would open an access route to this region allowing possible interactions between a sulfur containing moiety and the [2Fe-2S] binding region.

Unit	Binding interactions	Previous evidence
[4Fe-4S]	Cys53, 57 and 60 N/O bidentate with amino tail of SAM	Mutational studies. Cluster binding (84, 85) Spectroscopic studies (15)
[2Fe-2S]	Cys97, 128, 188 Arg260	Mutational studies. Vital for BS activity (89, 107)
SAM	[4Fe-4S], Arg173, Asn153, Asp155, Ile192, Tyr59	
DTB	Asn222, 151 and 153, Thr292 and 293, Adenine ring?	
Arg260	Ser43, 218, 283 and Arg95	

Figure 1.21a – Major binding interactions within the active site of BS crystal structure (15).

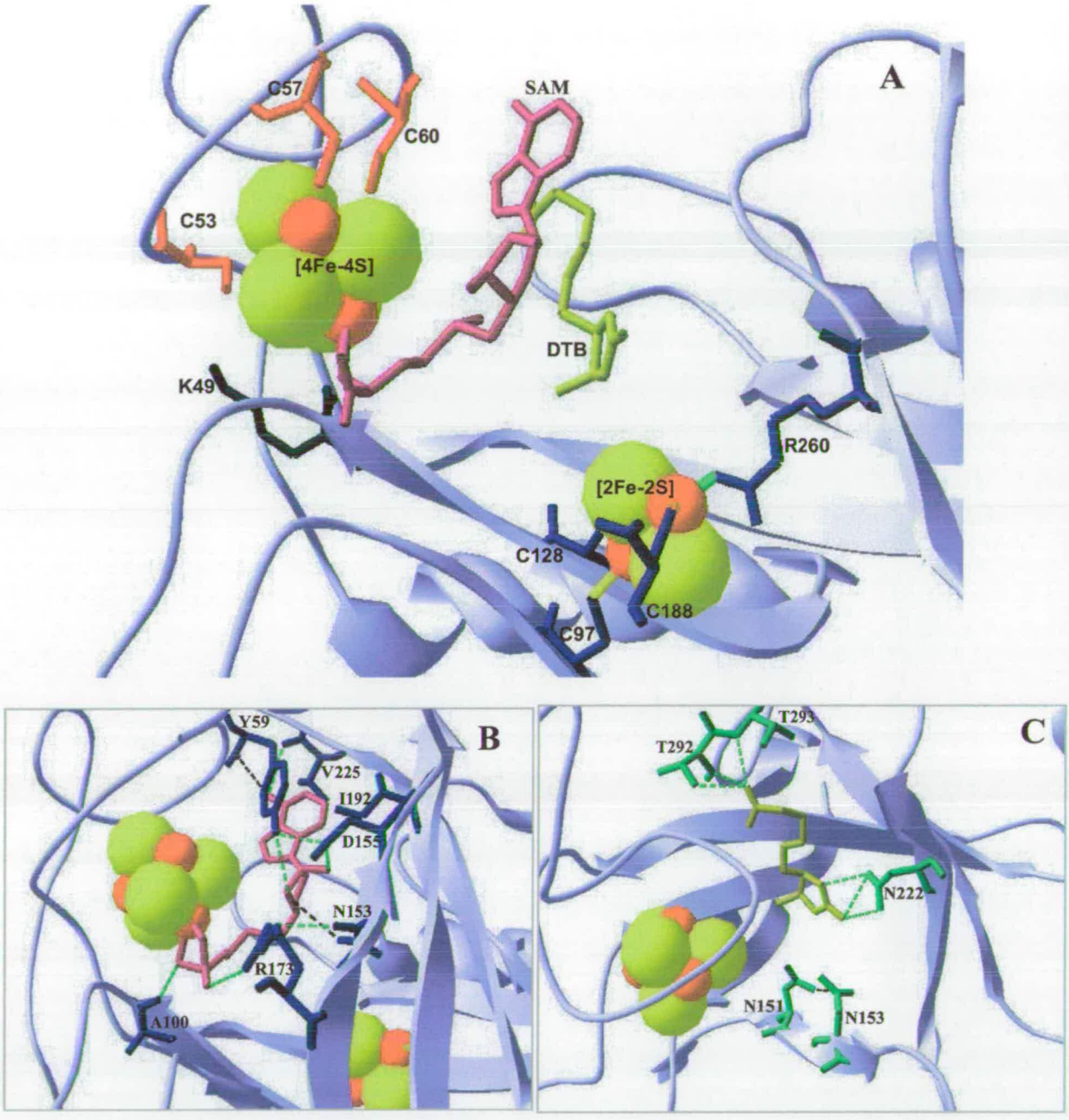


Figure 1.21b – X-Ray crystal structure of active site of *E. coli* biotin synthase (PDB ID - 1R30) showing clusters as spacefill, iron in red and sulfur in yellow. SAM is shown in purple and DTB in yellow. **A** – Displays both clusters and ligands thought to contact these. [4Fe-4S] cluster ligands shown in red (C53,57,60), final ligand is N/O chelate from SAM. [2Fe-2S] cluster ligands shown in blue (C97,128,188 and R260), residues contacting R260 (S43,218,283 and R95) omitted for clarity. Also shown is proposed PLP modified residue K49 (olive green). **B** – Shows ligands contacting SAM in dark blue (Y59, A100, N153, D155, R173, I192 and V225) with hydrogen bonds dotted green. Both clusters are also shown. **C** – Shows ligands thought to contact DTB and displayed in light green (N151,153,222 and T292,293) with hydrogen bonds in dotted green. The [2Fe-2S] cluster has been omitted for clarity (16).

1.4 – [Fe-S] cluster utilizing enzymes

Biotin synthase is the best characterised member of the growing family of radical-SAM, [Fe-S] cluster-dependent enzymes which catalyse diverse chemistry. Four other members which have been the subject of numerous studies are lysine 2,3-aminomutase (LAM), pyruvate formate lyase activating enzyme (PFL-AE), anaerobic ribonucleotide reductase activating enzyme (ARR-AE) and lipoic acid synthase (LIPA). Although all of these enzymes share in common with biotin synthase the requirement for [Fe-S] clusters and SAM subtle differences are apparent in their specific mechanism and less subtle differences in function as described below.

The enzyme lysine 2,3-aminomutase (LAM) has been purified from *Clostridium substerminale* and has been implicated in the inter-conversion of L-lysine and L-β-lysine. The active form of LAM has an absolute requirement for an [Fe-S] cluster, SAM and PLP. The protein forms a homohexamer with subunit weight of 47kDa producing an active complex of 285 kDa. The [Fe-S] cluster of LAM can be bound in the complex in three possible formations, but only one is thought to interact with SAM (112). It has been proposed that LAM and biotin synthase share a similar mechanistic pathway in which SAM binds directly to the cluster and is cleaved to form methionine and the 5'-deoxyadonosine radical (DOA$^{\bullet}$). In the case of LAM the resultant DOA$^{\bullet}$ radical is thought to abstract a hydrogen from L-lysine resulting in the formation of a lysine radical. This proposed lysine radical could isomerise, with addition of a hydrogen atom, to form the L-β-lysine product (112).

Pyruvate formate lyase (PFL) is a homodimer with monomer molecular weight of 85kDa the tertiary structure of which (113, 114) reveals a single stable glycyl radical (Gly734) which has been shown to be required for catalytic activity (115). The PFL dimer forms a $\alpha_2\beta$ complex with a second monomeric protein, pyruvate formate lyase activating enzyme (PFL-AE) a 38kDa enzyme containing a [4Fe-4S] cluster. In common with biotin synthase the [$(PFL)_2$:PFL-AE] complex also requires SAM, DTT, NADPH and the FLAV/FRED electron transfer complex to exhibit enzymatic activity (116, 117). The activation of this $\alpha_2\beta$ complex is thought to be initiated via a similar reduction cascade described for biotin synthase (NADPH → Flav → Fred →

[Fe-S]). The bound SAM molecule is cleaved to form the DOA$^{\bullet}$ radical which is then available to abstract a hydrogen atom from the α-carbon of the conserved Gly734 forming the glycyl radical required (118, 119). Three conserved cysteine residues (Cys 29, 33 and 36) have been proved to be essential for activity and also the site of cluster binding corresponding to the cys-box motif -C-X_3-C-X_2-C- (118).

Anaerobic ribonucleotide reductase (ARR) also forms a complex with its activation enzyme, anaerobic ribonucleotide reductase-activation enzyme (ARR-AE). ARR itself was found to be a homodimer and upon complexation with ARR-AE forms a $\alpha_2\beta_2$ complex which demonstrates enzymatic activity. Anaerobic ribonucleotide reductase activating enzyme (ARR-AE) contains a [4Fe-4S] cluster which is located in the interface between the β monomers (120). The equimolar binding of SAM to this cluster initiates the cleavage to form methionine and the DOA$^{\bullet}$ radical. In a mechanism similar to PFL the DOA$^{\bullet}$ radical abstracts a hydrogen atom from a conserved glycine residue (Gly681) forming the activated ARR enzyme.

The lipoate synthase (LIPA) enzyme catalyses the insertion of 2 sulfur atoms into an octanoic acid group during the final step of lipoic acid synthesis as laid out in Figure 1.22. This insertion reaction catalysed by LIPA is comparable to the conversion of DTB to biotin by biotin synthase as both involve the production of two new C-S bonds (76, 121, 122). The LIPA protein has approximately 36% identity with BS and both proteins have a requirement of an [Fe-S] cluster and the mechanisms are SAM dependent. Exogenous lipoate is transferred to apo proteins by lipoate protein ligase (LplA) (123). In the 2-oxoacid dehydrogenases, the lipoyl modified domains are found at the termini of the E2 subunit. Two distinct pathways exist in *E. coli* for the formation of lipoyl modified dehydrogenase E2 domains, so forming lipoyl-E2. Both involve the formation of an octanoylated-E2 before its conversion to the lipoyl-E2 by LIPA. The first pathway uses octanoic acid as the substrate (derived from the fatty acid synthetic pathway) and this is covalently linked to a dehydrogenase enzyme subunit by the LplA protein (123). This second pathway requires the *lipB* gene product and this utilises the lipoyl(octanoyl)-ACP as the substrate to form the octanoylated-E2, which again is converted to lipoyl-E2 by LIPA (124, 125). A recent study on the thermostable lipoyl synthase for Sulfolobus solfataricus by Roach *et al.* has discovered that the LIPA protein is quite promiscuous with regard to its substrate

and they developed an assay for activity based on octanoylated tetra peptides (126). Although the exact mechanism for the activity of LIPA is not known, a study by Booker *et al.* has shown that both sulphur atoms are provided from a single LIPA protein using an in vitro assay system containing a mixture of ^{32}S and ^{34}S labelled LIPA (127).

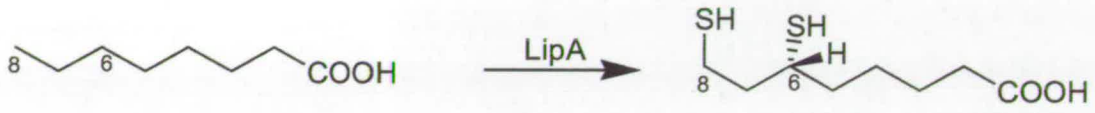


Figure 1.22 – Reaction catalysed by the *lip*A gene product, lipoic acid synthase, a [Fe-S] cluster containing member of the SAM-radical family. Reproduced from Fontecave *et.al.* (9).

1.5 - 3D Structures of Radical-SAM enzymes

Structures of three distinct SAM radical enzymes from the family of >600 representatives have been elucidated to date namely BS, HemeN and MoaA and a structural comparison is shown in Figure 1.23.

The HemeN enzyme catalyses the oxidative decarboxylation of two propionate groups of the A and B rings of coproporphyrinogen III to form coproporphyrinogen IX under anaerobic conditions in *E. coli.* In analogy with other members of the Radical SAM super family a mechanism has been proposed where initial abstraction of hydrogen from the beta position of the propionate side chain forms a radical, which is further oxidised by as yet unknown partner yields an allylic carbocation that facilitates the formation of the vinyl group (5).

The first step in Moco biosynthesis is catalysed by the SAM dependant [4Fe-4S] cluster MoaA and MoaC. This reaction involves the conversion of a guanosine derivative to a precursor Z which is an oxygen sensitive 6-alkyl pterin with a cyclic phosphate. As with other members of this family the reaction pathway involves a reductive cleavage of SAM by the [4Fe-4S] cluster(s) (14,128).

An interesting comparison of these 3 members of the SAM radical family was made by Schindelin *et al.* when they obtained the structure of MoaA as shown in Figure 1.23. All three structures show high structural similarity in the N-terminal domain of the protein which is responsible for [Fe-S] cluster and SAM binding but differ in the C-terminal domain of the protein, presumably to facilitate the binding of structurally diverse substrates and cofactors.

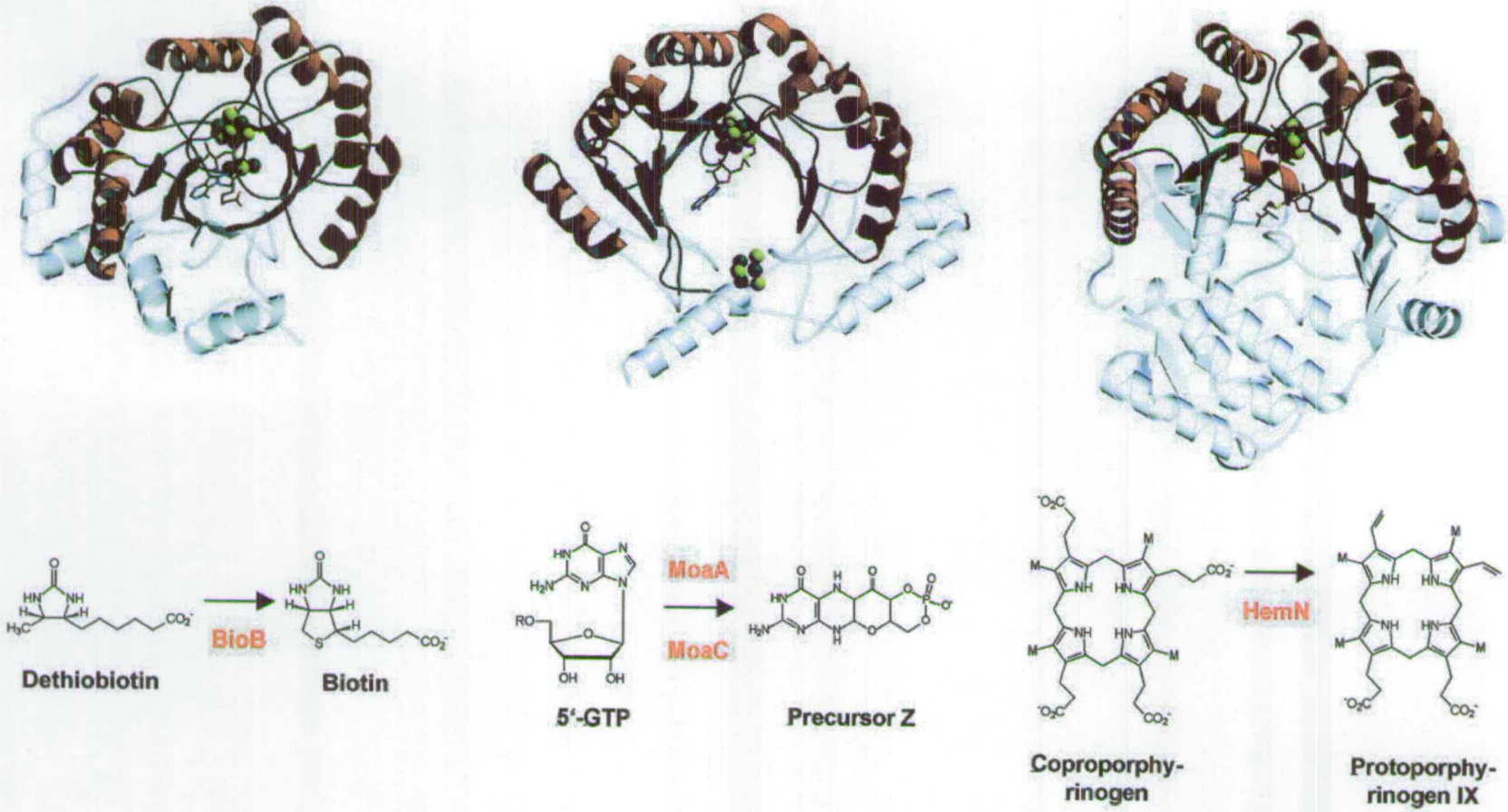


Figure 1.23 – A structural comparison of SAM-dependent enzymes BioB, MoaA and HemN and the reactions catalysed. The α-helicies and shown in red, β-sheets in blue and loops in grey. Structurally diverse regions are shown transparent. Reproduced from "Crystal structure of the S-adenosylmethionine dependant enzyme MoaA and its implications for molybdenum cofactor deficiency in humans" Petra Hanzelmann and Hermann Schindelin, PNAS vol 101, no. 35 (14).

1.6 – Bacterial Iron-Sulfur Cluster Formation

Identification of bacterial iron-sulfur clusters ("[Fe-S]" generic form) occurred over 40 years ago within the ferredoxin class of electron carrier proteins (129). Over 100 proteins are now known to contain [Fe-S] clusters and their structural, functional and mechanistic diversity is remarkable having influences in controlling protein structure, environmental sensing, modulation of gene expression and formation of reaction intermediates (130, 131). This diversity is almost certainly due to the chemical versatility of both iron and sulfur which in turn has led to suggestions that prebiotic iron-sulfur complexes may have played an influential role in the emergence of life (131, 132, 3). Although the most common forms known in biology are [2Fe-2S] and [4Fe-4S] many others have been identified with much higher nuclearity such as the [7Fe-9S-Mo] core of the nitrogenase MoFe protein (133). Studies by Holm and colleagues have shown that all clusters can be formed from reductive coupling of [2Fe-2S] units (134). This work demonstrated that clusters can self assemble, however it is unlikely that this spontaneous assembly occurs within the intracellular environment. The required components for this self assembly, $Fe^{2+/3+}$ and S^{2-}, are metabolic poisons and are therefore likely to be metabolised or detoxified before larger clusters can form. More likely is the use of protein scaffolds to form the [2Fe-2S] unit building blocks. Analysis of the *Azotobacter vinelandii* genome revealed a region of [Fe-S] cluster biosynthetic genes. The genes within this region became known as the iron sulfur cluster (*isc*) gene cluster and the predicted biological functions of gene products are shown in Figure 1.24. They include proteins with cysteine desulfurase activity, DNA binding motifs, electron transfer activity and chaperone function.

Subsequently, regions showing high sequence conservation to this *isc* gene cluster were found throughout numerous bacterial genomes. Their general involvement in synthesis of [Fe-S] clusters has been well established using genetic experiments in *E. coli* (135, 136).

Although it is generally accepted that the role of the isc proteins is to mobilise Fe and S in the cell, as yet, the specific function of every individual gene and their dynamic interactions as a whole are unclear.

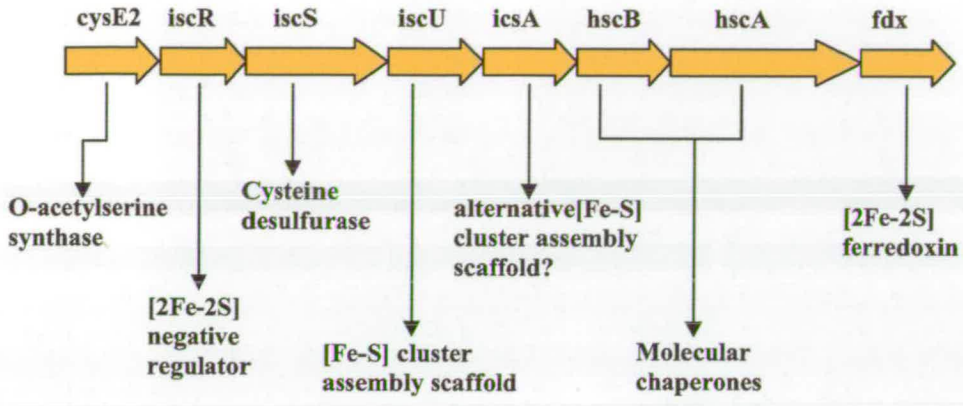


Figure 1.24 – Region of *A. vinelandii* genome encoding the iron sulfur cluster biosynthetic genes, *isc* operon (3).

The CysE2 gene product is an *O*-acetylserine synthase and this activity is known to be the rate limiting step in cysteine biosynthesis. It is assumed that the CysE2 protein facilitates the build up of the cysteine pool for downstream desulfurase activity (137). The *isc*R gene product is recognised as the protein responsible for regulation of the transcription of the *isc* machinery within bacterial cells. The mature iscR protein contains a weakly bound [2Fe-2S] cluster which can initiate a negative feedback on the operon. The weak cluster binding ensures that little or no competition for the products occurs (138). The product of *isc*S gene has been shown to be necessary for the mobilisation of sulfur for biotin, thiamine, thiolated tRNA and Mo cofactor maturation (139) and shows high sequence similarity to the cysteine desulfurase NifS from the nitrogen fixation pathway. The NifS gene product is located in a Nif-specific region of the genome and is involved in formation of [Fe-S] clusters specifically for the nitrogen fixation machinery (discussed in detail, Section 1.6). The mechanism of NifS is thought to proceed from the PLP dependant

desulfurisation of cysteine and formation of a NifS bound persulfide. The persufide is presumed to be the activated form of sulfur required by the scaffold protein NifU for [Fe-S] cluster assembly (140). General consensus suggests that the IscS and IscU proteins have similar functions to there equivalent Nif-specific genes (NifS and NifU) and although no unambiguous function has been identified for the iscA gene product it is accepted that it may function as the alternative scaffold protein (141, 142). The presence of the HscA and HscB molecular chaperones within this cluster was unexpected however it has been shown by Hoff *et al.* that these well studied chaperone proteins interact with the IscU protein (143). These chaperones have been implicated in either cluster assembly on the IscU or in the transfer of an assembled cluster to the target bio-molecule (3). The final gene in the cluster is the *fdx* gene encoding the ferredoxin protein that contains a permanent [2Fe-2S] cluster whose role within the *isc* machinery is not fully understood. Suggested functions include a site for transient [2Fe-2S] unit assembly into higher nuclearity clusters and a more recognisable role of involvement in the acquisition of Fe (3). The full understanding of exactly how the clusters are formed is of great interest throughout biochemistry because of the variety of cellular mechanisms that are implicitly linked to the [Fe-S] cluster reduction-oxidation cycle and therefore the formation of these within the cell needs to be better understood.

1.7 – Cysteine Desulfurase Enzymes

Studies by Dean *et al.* have greatly enhanced the understanding of sulfur trafficking within bacteria. It has been shown that the *nif* gene cluster in *A. vinelandii* is responsible for the construction of iron-sulfur clusters specifically for the nitrogen fixation machinery (144). The purified *nif*S gene product has been found to be a PLP dependant cysteine desulfurase, catalysing the conversion of L-cysteine into L-alanine and a protein bound persulfide. The persulfide is subsequently used in [Fe-S] cluster assembly (145, 6). The *E. coli* homolog of the nifS protein is IscS which demonstrates 40% amino acid sequence identity to *A. vinelandii* NifS.

Two other proteins have been identified in *E. coli* which display significant homology to NifS, cysteine sulfinate desulfinase (CSD) and CsdB or SufS, with sequence identity of 24%. These are presumed to represent a different class of cysteine desulfurase enzymes (146). All three identified NifS like proteins from *E. coli* (IscS, CSD and CsdB) have been shown to catalyse the desulfuration of cysteine and deselenation of selenocysteine with widely varying efficiency (147, 148). CSD shows wide specificity for a number of cysteine analogs such as L-cysteine, sulfinic acid, L-selenocysteine and L-cysteine. In contrast IscS and CsdB show high specificity towards one reagent, L-cysteine and L-selenocysteine respectively. The mechanism of activity for both cysteine desulfurase and selenocysteine deselenation, shown in Figure 1.25, is initiated by the conversion of the protein-PLP species, bound via a conserved lysine residue, to a substrate-PLP Schiff base upon introduction of substrate. Within the active site the conserved lysine residue which binds PLP could act as base to abstract a hydrogen and then turn H-donor for the subsequent step. The sulfur is thought to be abstracted by a conserved cysteine residue, thus forming the protein bound persulfide and L-alanine, with the reformation of the protein-PLP species (6). Recently, the crystal structure of *E. coli* IscS protein has been solved by Vickery *et al.* to a resolution of 2.1Å. The structure displays a homodimeric form with specific residues from the adjacent monomer protruding into its partner, shown in Figure 1.26. The conserved cysteine residue which is required for catalysis is Cys328 and it may have a dual role, where it also acts as the "arm" to deliver the persulfide to a partner scaffold protein,

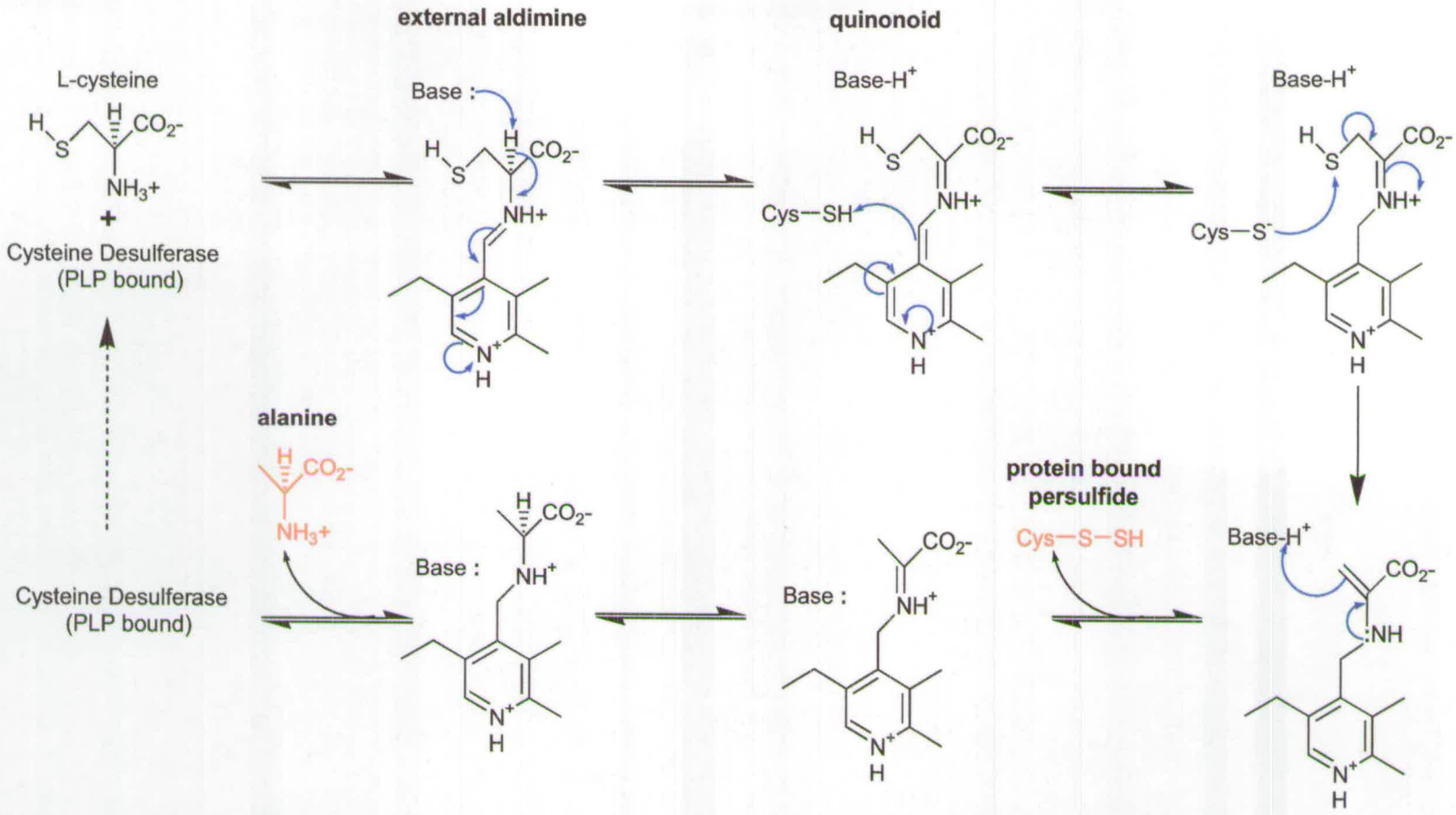


Figure 1.25 – Proposed mechanism of desulfuration of cysteine by NifS like enzymes, IscS, CSD and CsdB/SufS (6).

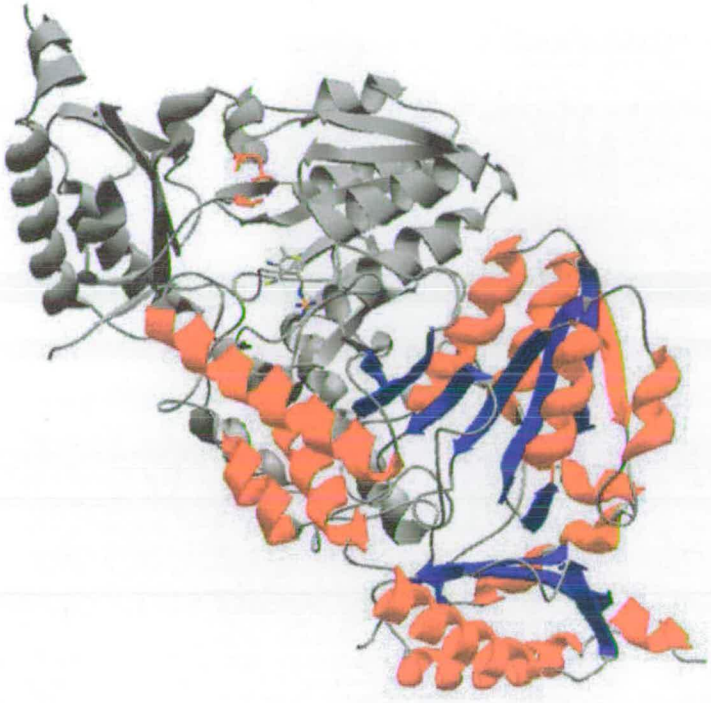


Figure 1.26 – The X-Ray crystal structure of *E. coli* IscS (1P3W). Dimer shown, with monomerA coloured β sheets blue and α helices red. MonomerB (grey) shows residues Ala237 and Glu334 highlighted in red showing location of ends of disordered loop containing conserved Cys328 (13).

which is presumed to be IscU/A (6, 149). The flexible nature of this "arm" containing Cys328 is emphasised by the location of the residue on a surface exposed disordered loop in the crystal structure. The termini of this disordered region lie at Ala327 and Glu334 (A^{327} and E^{334} residues highlighted in Figure 1.26). The assumption is that although this essential residue is not resolved in the structure its N-terminal neighbour is, Ala327, thus Cys328 must lie approximatly17Å from the active site and bound PLP. Modelling studies have shown that the extended loop could rotate to bring the γ-sulfur atom of Cys328 to within 3Å of the PLP, suggesting that large conformational changes occur during the catalysis process (13).

The binding pocket of *E. coli* IscS shows a large similarity to the NifS. The PLP molecule sits at the base of a binding pocket which in turn is near the surface of the protein molecule (see Figure 1.26 and 1.27). The pocket is highly solvated and lined

with several charged or polar amino acids including His104, Lys105, Asn155, Asp156, Tyr337 and Arg354. These residues are shown as space fill in Figure 1.27 and it is important to note the imidazole ring of His104 which is positioned directly below the pyridine ring of PLP. The PLP molecule is anchored into site via the conserved lysine residue, Lys206, via an internal aldimine Schiff base linkage as well as a number of polar and non-polar interactions. Hydrogen bonds (dashed green in Figure 1.27) are formed between phosphate oxygen's of PLP and the side chains of Thr76, Ser203 and His205 of that subunit one and Thr243 of the adjacent subunit (shown in *italics*). Additional hydrogen bonding occurs between the phenolate oxygen of PLP and Gln183 and between the pyridine N1 of PLP and Asp180. The structure is further stabilised by the hydrogen bonding of Asp180 to side chains of Asp79, Thr182 and Ser201 (not shown) (13).

One criticism of the proposal that BS is a PLP-dependant cysteine desulferase is that it lacks this typical PLP binding fold and that the lysine is surface exposed. However, all PLP-dependent enzymes undergo large conformational changes during their catalytic cycle and we must be careful not to rely on a single structure of [4Fe-4S], [2Fe-2S], DTB, SAM bound to BS. It is clear that more crystal structures of the protein in various stages of its reaction cycle (e.g. in the apo-form, and with methionine and biotin bound) will be required to fully understand the mechanism of DTB to biotin conversion. Certainly addition of PLP during the crystallisation process may provide important insight.

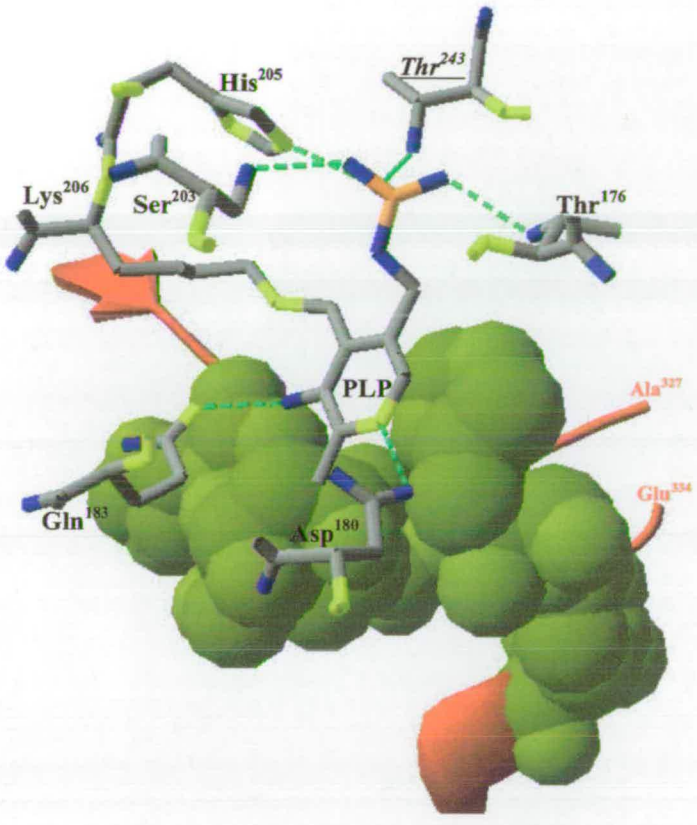


Figure 1.27 – Structure of the active site of *E. coli* IscS (1P3W). PLP linked via Schiff base to the Lys206 anchor and hydrogen bonding residues shown as labelled. Atoms coded - Carbon=Grey, Oxygen=Blue, Nitrogen=Yellow, Phosphorous=Orange. Charged and polar residues at bottom of binding pocket shown as space fill in dark yellow His104, Lys105, Asn155, Asp156, Tyr337 and Arg354. Termini of missing dis-ordered loop containing Cys327 shown as ribbon in red.

1.8 – Mass Spectrometry based proteomics

1.8.1 - Introduction

Over the last 10 years the term proteomics has become synonymous with mass spectrometry based analysis of biological samples. In general the proteomics discipline can be described as "...dealing with the large scale determination of gene and cellular function directly, at the protein level (150)." And encompasses a diverse range of techniques such as cell imaging, gene array technologies, genetic readout experiments such as the Yeast two-hybrid assay as well as mass spectrometry.

The development of new 'soft' ionisation techniques in the late 1980's, namely electrospray ionisation (ESI) (151) and Matrix assisted laser desorption ionisation (MALDI) (152), led directly to further application of mass spectrometry. Previous to these new ionisation techniques analysis was limited to small themostable compounds only. ESI and MALDI enabled a much greater diversity of compounds, and pertinently for this thesis, led to the ability to generate gas phase biomolecules without excessive source fragmentation (153). These developments were rightly acknowledged with the 2002 Nobel prize for chemistry.

These developments rapidly increased the power of mass spectrometry for research scientists and subsequently led to the development of new mass analysers and then to novel multistage MS/MS instruments (e.g. Q-ToF, ToF-ToF, QQQ etc).

Mass spectrometry based proteomics is most commonly used as a technique for the characterisation and/or identification of proteins and peptides derived from a biological source, regardless of the biological question being asked.

1.8.2 – Mass spectrometer basics

In the simplest terms a mass spectrometer consists of 3 portions, a source region where the ions are generated, an analyser region where ions can be sorted/filtered based on the mass to charge ratio (m/z) and a detector which can detect an arrival of an ion (150). The range of analysers currently on the market is huge and diverse. Here we will briefly describe the three instruments used during this PhD work and explain the analyser configurations.

Waters-Micromass Platform I

The Micromass Platform instrument is a mass spectrometer that contains an ESI source with a single quadrupole analyser. A schematic is shown in figure 1.28. The ions are generated in the ESI source region and passed to a quadrupole analyser before detection on the detector.

The quadrupole consists of four parallel rods. The rods have a circular cross-section and are equidistant from a central axis. A combination of DC and RF voltages are applied to the rods. Opposing rods have the same polarity (positive or negative); adjacent rods have opposite polarity. The combination of voltages separates ions on the basis of their mass-to-charge ratios (m/z). At a given set of voltages, only a single m/z value has a stable trajectory along the length of the rods allowing transmission to the detector (154). During data acquisition the DC voltage is continuously 'scanned' across voltages which allow the transfer of ions with in the required m/z range. At any given moment only an ion of a particular m/z will be passed through and thus detected, correlation between the time the ion was detected and the DC voltage applied provides the m/z and over the course of an experiment a picture is constructed of the component ions and there observed relative abundance.

Applied Biosystems Voyager-DE™ STR

The ABI Voyager-DE™ STR Biospectrometry™ Workstation is a MALDI-TOF (matrix-assisted laser desorption ionization time-of-flight) mass spectrometer that includes a reflector analyzer (TOF). A schematic is shown in Figure 1.28

Time-of-flight mass spectrometry works on the principle that if ions are accelerated with the same potential from a fixed point and time and are allowed to drift, the ions will separate according to their mass-to-charge ratios. Lighter ions drift more quickly to the detector and heavier ions drift more slowly. Ions are formed in a field-free region, and a high voltage pulse is applied after a predetermined time delay to accelerate the ions into the TOF region. A reflector acts as an electrostatic mirror and in principle, ions with higher kinetic energy penetrate further into the reflector and therefore take longer to emerge, thus decreasing the spread of ion velocities as they reach the detector (154). The reflector improves resolution by narrowing the

spread of velocities for ions of the same m/z. All generated ions are pushed into the TOF region at once and the time of arrival at the detector after the push correlates to the m/z.

Waters-Micromass QTOF

The QTOF (quadrupole time-of-flight) instrument is a hybrid instrument and contains both a quadrupole and a time of flight analyser, as shown in Figure 1.28, and is equipped with an electrospray source.

The instrument can be used in both MS mode (RF voltage only to quadrupole allowing passage of all ions) where m/z is measured in the TOF analyser and in what is known as MS/MS mode. In MS/MS the quadrupole DC voltage is set to allow transfer of only a specific m/z through, after this the filtered ion(s) are passed into the argon containing collision cell before m/z analysis in the TOF region. By increasing the energy supplied to the ions before entry into the collision cell the ions can be fragmented by a process called collision induced dissociation (CID) (154).

With this process it is possible to select an ion from a mixture, a single peptide ion from a digestion for example, and obtain not only the m/z of the peptide but m/z of fragments which may allow structural information to be elucidated. In the case of peptides specific fragmentation pathways are favoured, predominantly at the amine bond, allowing a subset of the peptide sequence to be elucidated.

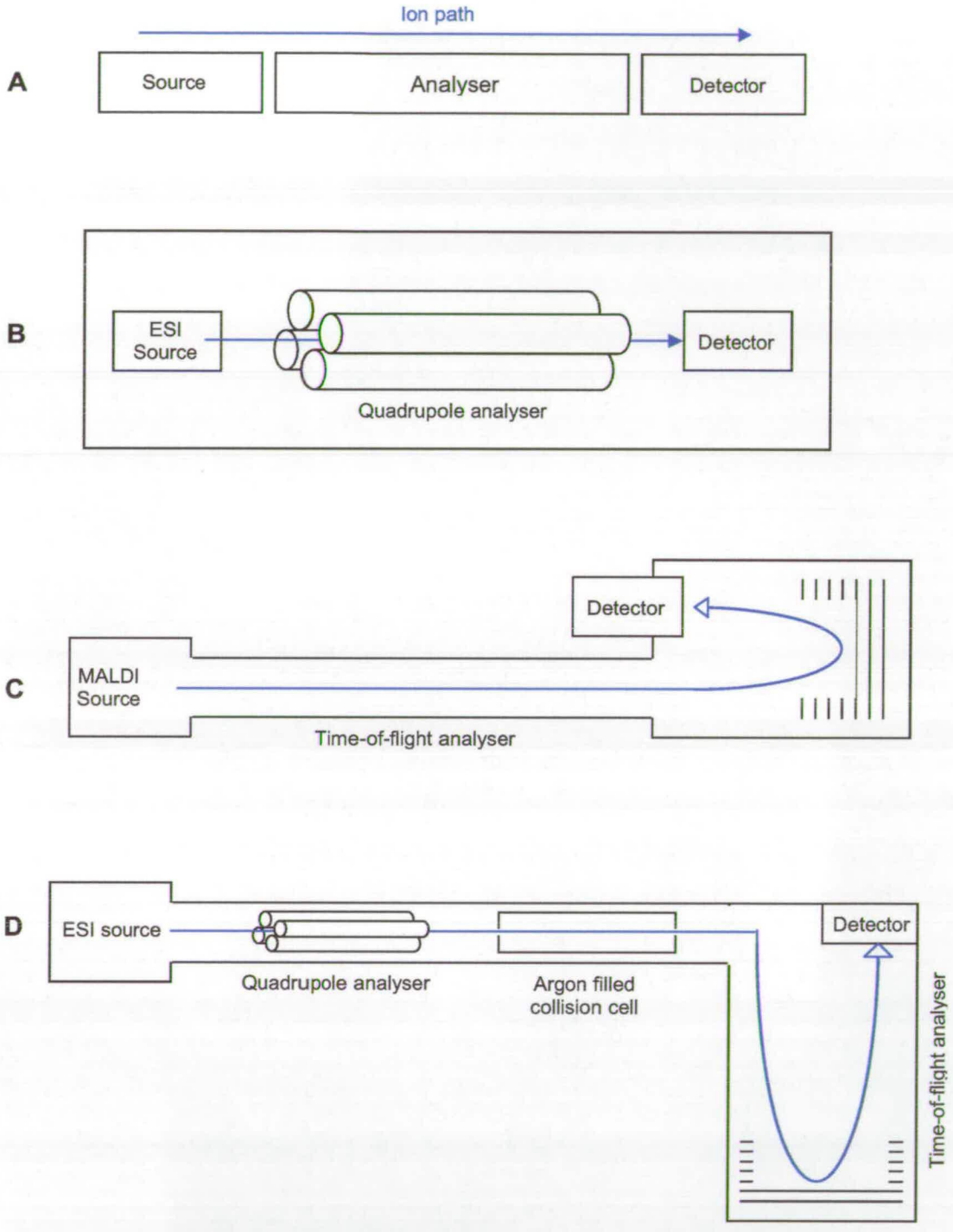


Figure 1.28 – General layout diagrams of mass spectrometry analysers. **A** – general MS; **B** – Micromass Platform I; **C** – Applied Biosystems Voyager DE^{TM}STR; **D** – Micromass Q-TOF

1.8.3 – Mass spectrometry of biomolecules

The basic approaches to most proteomic applications can be separated into 2 catagories. The analysis of 'intact' and the analysis of 'proteolytically digested' proteins, both of which can be extremely valuable in addressing specific problems. The application of the correct analysis can be both complementary and also provide specific information which is exclusive to that method.

The analysis of intact proteins has been used to simply measure the mass of a protein, in the case of recombinant proteins confirm global amino acid composition, and more recently to look for possible post-translational modifications such as phosphorylation and glycosylations. In this instance single stage mass spectrometers are used such as single quadrupoles and TOF instruments (153, 155). However given the nature of protein ionisation the analysis using an electrospray source, compared to MALDI, provides much more convincing and accurate mass measurement. This is mainly due to the multiple charges that typically are observed when a protein is ionised by electrospray (151, 11). In this thesis all intact protein mass spectrometry analysis has been done on the Waters-Micromass Platform I instrument coupled to a Waters 2795 HPLC and the results are discussed in Section 3.4.

In contrast, the analysis of digested proteins can be performed using a variety of instrumentation, separation methods and sample preparation. Two-dimensional gel electrophoresis (2DE), itself a common technique for comparative studies for over 25 years was the first common entry point to MS analysis (150). The traditional method for protein regulation analysis done using 2DE gives quantative information based on changing protein spot patterns within the gel. However the subsequent identification was extremely time consuming until mass spectrometry provided a quick and reasonably reliable resolution.

The general approach is to excise protein spot of interest, perform an enzymatic in-gel digestion of the protein before extraction of resultant peptides. These peptides are analysed using MALDI-TOF to produce a list of m/z ions which are characteristic of the protein (154, 150). The subsequent identification is achieved by a matching these ions with those predicted from an in-silico digestion of each protein from a database or subset. This method is known as peptide mass

fingerprinting (PMF) and is still the most common place method for protein identification (see Figure 1.29) (156). This technique has developed and now purification of proteins by other means (HPLC, CE etc.) are now commonplace as sample preparation for MALDI-TOF PMF analysis for protein identification.

Identification by PMF is based on m/z only and the technique is mainly applicable to purified proteins only. Extra confidence in protein identification can be gained by using MS/MS capable instruments (e.g. QTOF) to obtain not only peptide masses but also a partial or full peptide sequence. This enables database searching/matching based not only on m/z of peptides but also on their sequence. This can greatly enhance the confidence in protein identification even when only one or two peptides are observed (see Figure 1.29).

The need to identify proteins from only a few peptides is becoming more important as the desire to investigate increasingly complex samples is growing and integrated separations technologies are enhancing separation and hence increasing the ability to identify more components from highly complex samples. Identification of a protein based on a few peptides can also allow more information to be extracted from the analysis. For example, a peptide may have a post translational modification which alters the m/z and so will not be recognised from the database during a standard PMF search. However, the MS/MS spectra may provide the amino acid sequence of the peptide and hence the identity. Given the mass difference of expected to the observed it may be possible to putatively identify not only the modification but also the site of modification (see Figure 1.29) (155).

The field of proteomics is constantly developing and further advances in sample preparation, instrumentation and informatics, all of which are an integral part of proteomics, continue to push forward what it is possible to routinely achieve.

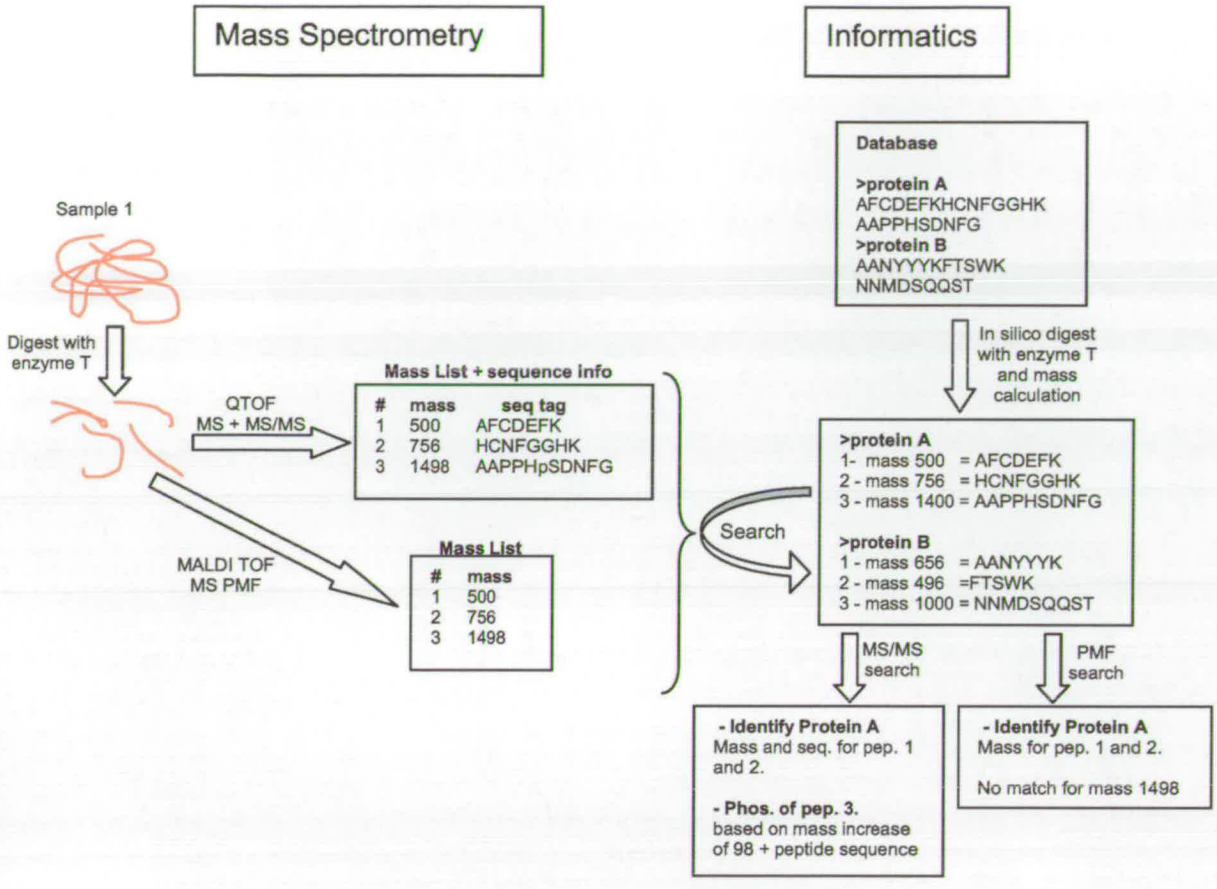


Figure 1.29 – General proteomics workflow showing both mass spectrometry and informatics input. Highlights the difference inceased confidence (peptide sequence information) and additional information (possible modifications) gained from MS/MS acquisitions.

Chapter Two: Materials and Methods

2.1 – General Reagents

All chemicals were purchased from Sigma-Aldrich, Amersham Pharmacia, Gibco BRL, New England Biolabs, Waters, or Invitrogen unless specifically stated. All reagents purchased were of appropriate quality for their use.

2.2 – Solutions and Buffers

Ampicillin – Stock solution of 100 mg/ml which was sterile filtered (0.45 μm) and was used at a final concentration of 100 μg/ml and stored at 4°C.

Kanamycin – Stock solution of 30 mg/ml which was sterile filtered (0.45 μm) and was used at final concentration of 30 μg/ml and stored at 4°C

Ethidium Bromide – A 10 mg/ml stock was made and used at final concentration of 5 μg/ml, stored at 4°C in a dark bottle.

IPTG – A 100 mM stock solution of Isopropyl β, D - thiogalactopyranoside was made and used at final concentration of 1 mM, filter sterilised and stored at 4°C.

Buffer A – Imidazole (5 mM), sodium chloride (500 mM), Tris-HCl (20 mM, pH 7.9)

Buffer B – Imidazole (1 M), sodium chloride (500 mM), Tris-HCl (20 mM, pH 7.9)

Buffer C – Tris-HCl (50 mM, pH 7.5)

Buffer D – Tris-HCl (50 mM, pH 7.5) + 2 M NaCl

SDS sample buffer – Tris-HCl (0.5 M, pH 6.8, 1 ml), Glycerol (2.0 ml), SDS (10% w/v, 1.6 ml), 2-β-mercaptoethanol (0.4 ml), bromophenol blue (0.05% w/v, 2.0 ml)

TAE – Tris-HCl (40 mM, pH 8.0), acetic acid (57.1 ml), EDTA (0.5 M, pH 8.0, 100 ml)

2D electrophoresis solutions and buffers

Cell lysis solution (stock)– 8 M Urea (19.2 g), 4% (w/v) CHAPS (1.6 g), 40 mM Tris base (0.194 g), double distilled water to 40 ml. Stored in 1 ml aliquots at minus 20°C. 5 µl of appropriate IPG buffer (Amersham Pharmacia) and 2 mg of DTT is added prior to use.

IPG strip rehydration buffer – 8 M Urea (12 g), 2% (w/v) CHAPS (0.5 g), Bromophenol blue trace (a few grains), Double distilled water to 25 ml. Stored in 2.5 ml aliquots at -20°C. 7 mg DTT and 50 µl of appropriate IPG buffer (from Amersham Pharmacia) are added prior to use.

SDS equilibration buffer – 50 mM Tris HCl (1.21 g) pH 8.8, 6 M Urea (72.07 g), 30% (v/v) Glycerol (69 ml), 2% (w/v) SDS (4.0 g), Bromophenol blue trace (a few grains), double distilled water to 200 ml. Stored in 10 ml aliquots at -20°C. Either 100 mg of DTT or 800 mg iodoacetamide is added to 10 ml prior to use.

Solutions for In-Gel digestion of protein–

100 mM ammonium bicarbonate

50 mM ammonium bicarbonate

20 mM ammonium bicarbonate

50% acetonitrile

5% formic acid in 50% acetonitrile

200 µg trypsin in 200 µl 50 mM ammonium bicarbonate

2 µg AspN in 200 µl 50 mM ammonium bicarbonate

10 mM DTT, 0.02% EDTA in 100 mM ammonium bicarbonate

50 mM iodoacetamide in 100 mM ammonium bicarbonate

Solutions of ZipTip procedure –

Wash – H_2O + 0.1% trifluroacetic acid

Wet / Elution – Acetonitrile : H_2O (1:1) + 0.1% trifluroacetic acid

Matrix solutions for MALDI-MS –

CHCA – Saturated solution of α-cyano-4-hydroxycinnamic acid in Acetonitrile : H_2O (1:1) + 0.1% trifluroacetic acid.

Sinapinic - Saturated solution of sinapinic acid in Acetonitrile : H_2O (1:1) + 0.1% trifluroacetic acid.

Dithionite / EDTA for [Fe-S] cluster reduction – 100 times stock, 50 mM Tris-HCl, pH7.5, 5% glycerol, EDTA (480 mM). Degassed with N_2 for >2h, dithionite added to final concentration of 480 mM in anaerobic glovebox. Buffer used within 30 min of addition of dithionite.

2.3 – Media

Agar Plates – 15 g agar per litre was added to specified media, 1000 times' stock antibiotic of choice added and approximately 25 ml poured into sterile Petri dish, stored at 4°C. X-Gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) was added to agar at 1 μg/ml final concentration.

2YT – bacto-tryptone (16 g), bacto-yeast extract (10 g), sodium chloride (5 g) were dissolved in one litre of water and pH adjusted to 7.5, autoclaved (121°C, 15 min, 15 psi).

Luria Bertani (LB) – Tryptone (10 g), yeast extract (5 g), sodium chloride (10 g), in one litre of distilled water and pH adjusted to 7.5 with sodium hydroxide.

2.4 – Bacterial Cell Lines

Cell Line	Genotype	Reference
JM101	F'; traD36; lacI^q; Δ(lacZ); proA$^+$B$^+$ / supE thi; Δ(lac-proAB)	(157)
JM109	F'; traD36; lacI^q; Δ(lacZ); proA$^+$B$^+$ / e14- (McrA-);Δ (lac-proAB); Thi gyrA96 Nal^r); endA1; hsdR17 (r_k^- m_k^-); relA1; supE44; recA1	(157)
Top10 (Invitrogen)	(F$^-$; mcrA; Δ(mrr-hsdRMS-mcrBC) Φ80*lac* ZΔM15; Δ*lac* X74; deoR; *rec*A1; araD139; Δ(*ara-leu*)7697; *gal*U; *gal*K; *rps*L; *end*A1; *nup*G.	(158)
HMS174 (DE3) (Novogen)	F$^-$; recA1 (r_{k12}^- r_{k12}^+) ; Rif^r (DE3), a λ prophage carrying the T7 RNA polymerase gene	(159)
BL21 (DE3) (Novogen)	F$^-$; ompT LsdS$_B$ (r_B^- m_B^-) gal dcm(DE3)	(160)

2.5 – DNA Vectors and Plasmids

Vector used for initial cloning and storage of genes was pCR2.1 TOPO cloning vector, which contains a precut TA overhang for direct cloning of PCR products, an ampicillin and kanamycin resistance gene.

Plasmids used for expression of proteins were, pET16b (Phenotype : amp^r ; T7 promotor ; Factor Xa protease cleavage site), pET6H (Phenotype: amp^r ; T7 promotor : Factor Xa protease cleavage site ; N-terminal 6His sequence), pHisBS (His$_6$-biotin synthase), pC53/57S, pC53/60S, pC57/60S, pC53/57/60S, pK49Q, pK49R, pBS-N$_{term}$, (N$_{term}$ His$_6$-tagged BS mutants, Amp selection) and pBS-C$_{term}$, (deletion mutant BS with Kan selection).

2.6 – Oligonucleotide primers

The following oligonucleotide primers were custom made by Sigma-Genosys, restriction sites in bold, codon changes underlined, stop codon highlighted in grey, methionine start highlighted in green.

BioBfor – 5'-GGTAAAACC**ATG**GCTCACCGCCCA-3'
BioBrev – 5'-**GGATCC**TCATAATGCTGCCGC-3'
C53S – 5'-AAGACCGGAGCTTCCCCGGAA-3'
C57S – 5'-CCGGAAGATTCTAAATACTGCCCG-3'
C60S – 5'-AAATACTCTCCGCAAACGTCGCGC-3'
M13for – 5'- GTTGTGTGGAATTGTGAGCGG -3'
M13rev – 5'- CGCCAGGGTTTTCCCAGTCACG -3'
BSNterm_rev – 5'-**GGATCC**TCAATCATCGTTATCGGCAAG-3'
BSCterm_for – 5'-**CCATG**GTCGATGCCTTTGATTTTATT-3'

BioBfor contains *Nco*I restriction site for cloning and Met start codon, BioBrev contains *Bam*HI restriction site for cloning and TCA stop codon. BioBrev contains a *Bam*HI restriction site for cloning and a stop codon. C53S, C57S and C60S all contain 3 base pair mutations encoding a cysteine to serine amino acid sequence change. M13for and M13rev were using for sequencing of genes from PCR2.1 vector.

2.7 – Equipment

Denley BR401, Micro centrifuge Spectrofuge 16M and Sorvall RC-5B centrifuges were used for all centrifugations. Unicam UV/Vis spectrometer UV4, Labnet Thermolyne G10 gyrorotory shaker used for cell culture. Horizontal running frame (Berthesda Research Labs, Model H5) for Agarose Gel electrophoresis, vertical running frame (Bio Rad Labs), power packs (Anachem PSU 400/200, Bio Rad Power pac300), precast Nu-PAGE gel running apparatus (Invitrogen) for gel

based 1D protein electrophoresis. IPGphor (Amersham Pharmacia), Mutiphor II (Amersham Pharmacia) used for 2D gel based protein electrophoresis. Mass Spectrometry was performed on Micromass Platform1 equipped with online waters HPLC-2690, Voyager DE™ STR Mass Spectrometer (ABI) and Micromass Q-ToF1 with Nanospray and ESI sources.

2.8 – Transformation of Competent Cells

Competent cells are stored at -80°C, 25 µl aliquot of required cells are defrosted on ice. 2 µl of plasmid DNA is added to cells and incubated on ice for 10 min to 30 min, cells heat shocked at 42°C for 30 sec and returned to ice for 2 min. 250 µl of SOC media is added and shaken at 250 rpm and 37°C for 1 hour. 25-250 µl of cells are spread to dryness on selected agar plate (161).

2.9 – Plasmid DNA preparation

Cells transformed with required DNA, were grown up in 10 ml of 2YT broth with appropriate selective antibiotic. Cells harvested by centrifugation and DNA purified with Qiagen plasmid DNA Mini Prep Kit. The plasmid DNA stored in ddH_2O at -20°C.

2.10 – Restriction endonuclease digestion of DNA

Typically, the required amount of DNA (0.5 µg to 1 µg) was treated with required restriction endonuclease in appropriate buffer for 1 to 3 hours at 37°C (amount of enzyme and buffer as advised by manufacture). Blue/orange loading dye (10 times stock) added to incubation mixture prior to agarose gel electrophoresis (161).

2.11 – Agarose Gel Electrophoresis

The required amount of molecular biology grade agarose was added to 100 ml of TAE buffer and heated to 100°C for the agarose to dissolve (typically 1 g per 100 ml buffer produces a 1.0% Gel). The solution was cooled to ~50°C before 5 µl of stock ethidium bromide was added (final concentration of 5 µg/ml) and poured into mould and allowed to set at RT. The DNA fragments were visualized under UV light, and sizes gauged using standard markers Hyperladder I and IV (Bio-Line) (161).

After visualisation excised DNA bands were purified from gel using Qiaprep gel extraction kit (Qiagen).

2.12 – PCR amplification and mutagenesis of DNA

All PCR and sequencing reactions were performed in Perkin Elmer 480 thermal cycler. PCR reactions set up in 0.5 µl PCR tube as follows.

1 µl template DNA,

1 µl forward primer,

1 µl reverse primer,

39 µl ddH_2O,

2 Ready-to-Go PCR beads (Taq Polymerase; Amersham Biosciences).

The reaction was overlaid with mineral oil and cycled 30 times for 1min at 95°C, required extension time (1kbase/min) at 54°C. The PCR product subjected to Agarose gel Electrophoresis and purification as in Section 2.11. In mutagenic PCR reactions, the first PCR used as reverse Mega primer in subsequent PCR reaction (162).

2.13 – Automated DNA sequencing

Automated DNA sequencing was carried out on an ABI prism 377 DNA sequencer using the Sanger dideoxy chain termination method. The following was combined in a 0.5 ml PCR tube,

template DNA (2 µl),
primer (1 µl),
ddH2O (13 µl),
BigDye sequencing mix (4 µl, PE Biosystems).

The reaction was overlaid with mineral oil and incubated at 95°C for 30 sec, 54°C for 15 sec and 72°C for 4 min which was cycled 30 times. Sequencing chromatograms were analysed with Vector NTi - Contig Xpress software.

2.14 – Cloning into Plasmid Vectors

PCR products were cloned into pCR2.1 vector (Invitrogen) and then restriction endonuclease-digested DNA was cloned into complementary cut expression plasmid DNA such as pET plasmids (see Section 2.10). PCR products were purified by Agarose Gel Electrophoresis (see Section 2.11) and the following were combined in a 1 ml Eppendorf tube; purified insert DNA (3 µl), Salt solution (1 µl), pCR2.1 vector (1 µl). The ligation reaction was incubated at room temperature for 10 to 30 min, and 2-4 µl of the reaction used to transform Top10 *E. coli* cells and the transformation mixture spread on an X-Gal-LB-amp plate. Positive colonies were identified by Blue/White screening.
Restricted DNA (vector and insert) were combined in 1 ml tube as follows,

Insert DNA (14 ml),
vector DNA (2 ml),
T4 DNA ligase (2 ml),
10 times T4 DNA ligase buffer.

The ligation reaction was incubated at RT for > 16 hours and used to transform Top10 *E. coli* cells. Positive clones were identified by restriction digest analysis and agarose gel electrophoresis.

2.15 – 1D SDS Poly Acrylamide Gel Electrophoresis

The discontinuous buffer system of Laemmeli (163) was used to analyse proteins on a molecular weight basis via SDS poly acrylamide gel electrophoresis.
The NuPAGE system (Invitrogen) of precast gels were also used where appropriate.

2.16 – Small Scale Induction of Protein Expression

A freshly transformed single colony from agar plate was used to inoculate 2 ml of 2YT broth with appropriate selective antibiotic in a 25 ml universal tube. The culture was grown (37°C, 250 rpm shaking) to an OD_{600} = 1.0 (4 to 5 hours), 1 ml aliquot removed as control and placed in fresh tube and protein expression induced in remaining aliquot with the addition of 1 mM final concentration IPTG and grown for a further 3 hours. 250 µl of each were centrifuged (13k rpm, 2 min) and the supernatant was discarded, the pellet resuspended in SDS-sample buffer (250 µl), boiled for 5 min and 15 to 25 µl loaded onto SDS-PAGE (see Section 2.15).

2.17 – Preparative Induction of Protein Expression

Routinely 4 litre growths were performed in 8 one litre shake flasks as follows. A freshly transformed single colony from an agar plate was used to

inoculate 500 ml of 2YT broth with appropriate selective antibiotic in a 1 litre shake flask. The culture was incubated (37 °C, 250 rpm shaking) for approximately 16 hours. A 1 ml aliquot was removed as control, and protein expression was induced with the addition of 1 mM final concentration of IPTG and incubated (37 °C, 250 rpm shaking) for a further 3 hours. The cell pellet was harvested by multiple centrifugation steps (5k rpm, 10 min) and the final wet cell paste weight calculated (typically 25 g from 4 litre growth).

2.18 – Preparation of Cell Free Extracts

The wet cell pellet was fully re-suspended in 3 times wet weight of appropriate chromatographic binding buffer and cell lysis performed by sonication at 4 °C for 15min (power on for 30 sec off for 30 sec). A Cell Free Extract (CFE) was obtained by two subsequent centrifugation steps (2 times 25 min at 20k rpm). CFE was 0.45 μm sterile filtered before addition to any chromatographic column.

2.19 – Purification of 6His-BS and histidine tagged BS mutants

Large scale preparation of 6His-BS was done using BL21(DE3) or HMS174(DE3) pHisBS transformants according to procedure set out in Section 2.17. 6HisBS CFE was prepared in BufferA (Section 2.2) ready for binding to HiTrap® chelating affinity column (Amersham-Pharmacia). The HiTrap® column was pre-charged with Ni^{2+} (Flow rate 2 ml/min at 4 °C, 3 CV ddH_2O, 5 CV 100 mM $NiSO_4$, 3 CV Buffer A) and CFE loaded onto the column at 4 °C with a flow rate of 1 ml/min. The column was washed with 5 CV of Buffer A and protein eluted with a 50 CV gradient from 100% Buffer A to 100% Buffer B. Fractions containing 6HisBS protein typically eluted in the region of 200 mM imidazole and purity was judged by SDS-PAGE and concentration by OD_{280}. Fractions containing pure 6His tagged protein were immediately dialysed vs. Buffer C and aliquots stored at -20 °C.

6HisBS mutants were also produced and purified using the same method as described above.

2.20 – Anaerobic preparation of apo-6HisBS

Approximately 40 mg of purified 6HisBS was concentrated to 400 μM using a 30kDa c.o. centrifugal concentrator (Vivascience). The concentrated protein and 2 aliquots (25 ml and 3.5 ml) of Buffer C (50 mM Tris-HCl, pH7.5) were deoxygenated by bubbling with N_2 for > 2 hours. EDTA/Dithionite buffer was also prepared in parallel (See Section 2.2). All buffers were transferred to N_2 glovebox and 25 μl of 100 times EDTA/Dithionite buffer added and incubated in glovebox at RT for 16-20 hours.

Reduced apo6HisBS was desalted and buffer exchanged using a PD10 gelfiltration column (Amersham-Pharmacia), 25 ml deoxygenated Buffer C used to equilibrate column, 2.5 ml protein sample loaded and eluted with 2.5 ml deoxygenated Buffer C. Sample removed from Glovebox and stored at -20°C (68, 164).

2.21 – PLP loading of 6HisBS and reduction of internal aldimine

Aliquots of 6HisBs (2.5 ml protein, approx 25 μM) were freshly loaded with PLP by dialysis verses two changes of 3 litres of Buffer C + 100 μM PLP for 2 hours each. Unbound PLP was removed by gelfiltration using disposable PD10 GF column (Amersham-Pharmacia, 25 ml equilibration, 2.5 ml sample and eluted with 3.5 ml) using Buffer C without PLP. If required, reduction of the protein-PLP internal aldimine was performed with DTT, final concentration 10 mM.

2.22 – Analytical Native Protein Chromatography

All native protein chromatography experiments (Gel Filtration and Ion Exchange) were run on AKTA Basic (Amersham-Pharmacia) at RT and concurrently

monitored at 280 nm, 330 nm and 425 nm wavelengths. Identity of elution species were confirmed by SDS-PAGE mass spectrometry. Columns used were MonoQ (anion exchange) and Superdex 200 (gel filtration) and were purchased from Amersham-Pharmacia.

2.23 – Two Dimensional SDS-Poly Acrylamide Gel Electrophoresis

2.23.1 – Cell Extract preparation

Typically 10^8 cells (1 ml of culture OD_{600} = 1.0) were pelleted by centrifugation (13k rpm, 5 min) and were resuspended in 1 ml of Cell Lysis buffer (see Section 2.2) vortexed vigorously for 20 min followed by sonication in bath for 5 min. The Cell Extract was cleared by centrifugation (14k rpm, 30 min) and the supernatant removed and stored at -20 °C or -80 °C.

2.23.2 – Immobilized pH Gradient Isoelectric Focusing (1D)

The cell extract defrosted on ice and 20 µl added to 180 µl of IPG strip rehydration solution (See Section 2.2, with appropriate IPG ampholyte mix and DTT added) and mixed by inversion. The entire 200 µl was loaded to IEF running coffin and dehydrated strip placed gel side down on top ensuring no air bubbles are present and gel is in contact with electrodes. Sample and strip are covered with Dry Strip cover fluid and rehydration/IEF protocol performed using IPGphore at 20 °C.

Step	Time - h	Voltage – V
Rehydration	12	30
IEF – step 1	1	150
– step 2	1	300
– step 3	1	600
– step 4	1	1200
– step 5	1	2400
– step 6	4	8000
	Total =	37,000 Vh

Strips were equilibrated for SDS-PAGE immediately or stored at -20°C for subsequent use.

2.23.3 – SDS-PAGE of Isoelectric Focused Cell Extract (2D)

Focused IPG strips were incubated in 25 ml SDS equilibration buffer firstly with DTT added, secondly with iodoacetamide and finally with no addition for 10 min each. Equilibrated strips were placed gel side down onto precast, 0.5 mm Excel Gel on MultiPhore (Amersham-Pharmacia) with gel-buffer strips and run for 15 min at 30 mA and 600 V (constant 20 °C) until the coomassie dye front had fully entered SDS gel. The strip was removed and gel run for approximately 2 hours at 50 mA, 600 V until coomassie dye front had run from bottom of gel. Backed gel removed from MultiPhore and fixed immediately with 10% acetic acid, 40% Ethanol, 50% ddH_2O to allow subsequent staining.

2.23.4 – Staining of 2Dimensional Gels

All gels were equilibrated for staining with 70% methanol 7% acetic acid 23% H_2O for 30 min. Colloidal coomassie staining was performed with GelCode (Pierce) as per manufacturers' instructions. When required silver staining was achieved using the Plus-One Silver Stain kit (Amersham-Pharmacia) utilising Mass Spec compatible modified protocol (No glutaraldehyde of formaldehyde).

2.23.5 – Analysis of 2D Gels

Gels were scanned and analysed using the Genomic Solutions "Analyzer HT Investigator" software and target protein spots identified.

2.24 – Enzymatic Digestion of Protein

2.24.1 – In Gel Digestion

The area of interest from the gel (1D band or 2D spot) was excised with a clean scalpel, placed in a clean sterile 1.5ml tube and dried in a GyroVap for 30 min. Gel pieces were treated as follows, solutions described in Section 2.2. Gel pieces were washed twice with 50% acetonitrile for 5 min followed by 30 min drying in

GyroVap. Samples were reduced with DTT/EDTA solution for 30 min at 56 °C followed by alkylation of cysteines with iodoacetamide solution for 30 min in dark. Samples were dried in GryoVap before addition of 5-10 µl of required endoproteinase (e.g. trypsin). After 15 min 50 µl 50 mM ammonium bicarbonate added and incubated for 16 hours at 37 °C. Digestion liquid was removed and dried in GyroVap and stored at -20 °C.

2.24.2 – Solution Digestion

Large scale total enzymatic digestion of protein samples was performed by precipitation of protein with 50% acetone 10% trichloroaceticacid (TCA) for 30 min at 4 °C; sample was centrifuged for 30min at 14k rpm and supernatant removed. The protein pellet was washed 4 times with -20 °C acetone and resuspended in 50 mM ammonium bicarbonate. Samples were reduced and cysteines alkylated with DTT and iodoacetamide if required followed by addition of endoproteinase at final concentration suggested by manufacturer. The digested protein solution stored at -20 °C.

2.25 – Limited Digestion of Native Protein

A protein solution (typically Tris-HCl, pH7.5) was incubated with endoproteinase for required amount of time and reaction stopped by addition of SDS-PAGE sample buffer and boiling for 5 min or protein precipitation by addition of 50% acetone + 10% TCA at 4°C for 1 hour. Samples were analysed by SDS-PAGE, HPLC RP chromatography and Mass Spectrometry.

2.26 – UV-Visible Spectroscopy

UV-Visible spectra were measured on a Unicam UV/Vis spectrometer UV4 fitted with diode array using 0.5 ml quartz cuvettes and scans were performed between 250 nm and 600 nm.

2.27 – HPLC Reverse Phase Chromatography

All reverse phase chromatography was performed on Gilson 306 HPLC, mobile phase was suitable grade H_2O + 0.1% trifluroacetic acid (Buffer X) and acetonitrile (AcN) + 0.1% trifluroacetic acid (Buffer Y) at a flow rate of 0.5 ml/min. Columns were equilibrated with Buffer X and after sample loading 2.5 ml Buffer X used to wash unbound species from column before a 5 to 95% Buffer Y gradient. Columns used were Luna 5 µm C18(2) 250 mm x 4.6 mm, Luna 5 µm C8(2) 150 mm x 5 mm and Jupiter 5 µm 300 Å C5 150 mm x 2mm and purchased from Phenomenex.

2.28 – Liquid Chromatography Mass Spectrometry (LC-MS)

Electrospray mass spectrometry (ESI-MS) was performed on a Micromass Platform I single quadrupole mass spectrometer equipped with an electrospray ion source. The system conditions were first optimised with Horse Heart Myoglobin a further optimised with repeat injections of protein under analysis. The spectrometer cone voltage was set to 70 volts and source temperature to 65°C. Protein samples were separated on a Waters HPLC 2690 with Phenomenex C5 reverse phase column directly connected to the spectrometer. The proteins were eluted from column with 0.01% TFA and a 5% to 95% acetonitrile gradient at a flow rate of 0.2 ml/min and OD_{280} measured. Total ion current was measured from 500 to 3000 m/z. The protein peak scans were combined and average molecular mass was determined from the protein ion envelope by the MaxEnt1 and Transform algorithms of the Mass Lynx software (Waters-Micromass).

2.29 – MALDI-ToF Mass Spectrometry

Samples for MALDI-ToF analysis were co-crystallised with Matrix solution (Alpha-Cyano-4-hydroxy cinnamic acid (CHCA) or Sinapinic acid (see Section 2.2), directly on MALDI plate. MALDI-ToF peptide mass fingerprinting (PMF) was performed on an Applied Biosystems Voyager DE-STR.

Peptide spectra were acquired using CHCA matrix and instrument in reflectron mode with delayed extraction and positive polarity. Spectra were acquired in the range of 500-3500 Da with the low mass gate set at 500 Da. Other variables were set as follows, Accelerating voltage = 20,000V, Grid voltage = 65%, Mirror voltage ratio = 1.12, Extraction delay time = 170 nsec. Typically 500 shots per spectra were taken and a minimum 1000 shots were taken for each sample combined an analysed using the DataExplorer software from Applied Biosystems.

Protein spectra were acquired using Sinapinic acid Matrix and instrument in linear mode with delayed extraction and positive polarity. Spectra were acquired in the range of 500-45,000 Da with the low mass gate set at 500 Da. Other variables were set as follows, Accelerating voltage = 25,000 V, Grid voltage = 96%, Guide wire = 0.15%, Extraction delay time = 800 nsec. Typically 100 shots per spectra were taken and a minimum 400 shots were taken for each sample combined an analysed using the DataExplorer software from Applied Biosystems.

2.30 – Q-ToF Mass Spectrometry and Collision Induced Fragmentation MS/MS

2.30.1 – Nanospray Q-ToF MS^n

Samples were desalted prior to analysis using ZipTip (Millipore) according to manufacturers' instructions. Micromass Q-ToF I mass spectrometer was fitted with nanospray source and samples were sprayed in 50% acetonitrile + 0.1% TFA from Ag coated borosilicate capillary needles (Proxeon). Capillary voltage optimized for maximum signal intensity between 1000 and 2000 V with cone voltage set at 65 V. Argon was used as fragmentation gas which was set at 4 kV for all MS^1 acquisitions

and data was acquired from 500 to 3000 m/z with. When CID MS^2 was required the quadrupole RF set to transmit monoisotopic mass and 3 isotope peaks of species under study, fragmentation energy was increased to give optimum fragmentation pattern (up to 40 kV) and data acquired from 50-3000 m/z.

2.30.2 – Electrospray Q-ToF MS^n

The Micromass Q-ToF I mass spectrometer was fitted with ESI standard source and samples from HPLC were directly infused with a flow rate of 5 μl/min, capillary voltage was set to 3000 V and cone voltage to 65 V. Argon was used as fragmentation gas which was set at 4 kV for all MS^1 acquisitions and data was acquired from 500 to 3500 m/z. When CID MS^2 was required the quadrupole RF set to transmit monoisotopic and 3 isotope peaks of species under study, fragmentation gas increased to give optimum fragmentation pattern (up to 40kV) and data acquired from 50-3000 m/z.

Chapter Three: Results and Discussion

3.1 – Production of Biotin Synthase Mutant DNA

Previous work done by McIver *et al.* (87) has yielded a variety of vectors containing the *E. coli bio*B gene and the pET16b/BioB was chosen as template DNA for the site directed mutagenesis performed using the mega primer method (165). The BioB gene contains a natural *Nco*I (CCATGG) site at the 5' end incorporating the methionine start codon ATG. A *Bam*HI (GGATCC) site was engineered at the 3' end of the gene during the initial amplification and cloning reactions thus all subsequent cloning was done using *Nco*I and *Bam*HI restriction sites.

Previous work has demonstrated that the highly conserved 'cys box' motif, CX_3CX_2C, is the site of [Fe-S] cluster binding and single mutations at these conserved cysteine residues to serine significantly reduced the ability of the protein to bind these clusters (87, 86).

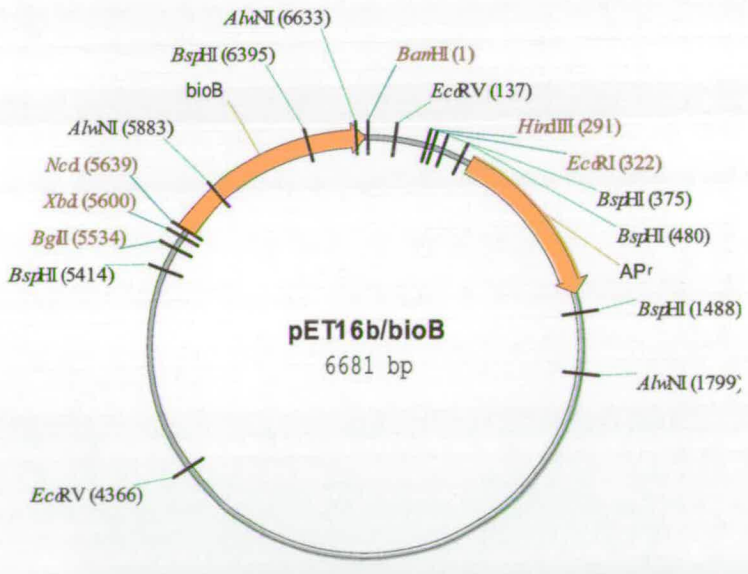


Figure 3.1 – Representation of the plasmid pET16b/bioB (6681bp) containing the 1041bp bioB gene cloned using the NcoI/BamHI cloning sites

```
bioB WT                 QVQVSTLLSIKTGACPEDCKYCPQSSRYKTGL
BioB C53S               QVQVSTLLSIKTGASPEDCKYCPQSSRYKTGL
bioB C57S               QVQVSTLLSIKTGACPEDSKYCPQSSRYKTGL
bioB C60S               QVQVSTLLSIKTGACPEDCKYSPQSSRYKTGL
bioB c53S/C57S          QVQVSTLLSIKTGASPEDSKYCPQSSRYKTGL
bioB C53S/C60S          QVQVSTLLSIKTGASPEDCKYSPQSSRYKTGL
bioB C57S/C60S          QVQVSTLLSIKTGACPEDSKYSPQSSRYKTGL
bioB C53S/C57S/C60S     QVQVSTLLSIKTGASPEDSKYSPQSSRYKTGL
bioB K49Q               QVQVSTLLSIQTGACPEDCKYCPQSSRYKTGL
bioB K49R               QVQVSTLLSIRTGACPEDCKYCPQSSRYKTGL
```

Figure 3.2 – Alignment of wild type biotin synthase with all constructed mutants. Residues as wild-type shown in red and mutated residues in blue. Residues mutated are lysine 49 (K49), cysteine residues 53, 57 and 60 (C53, C57 and C60).

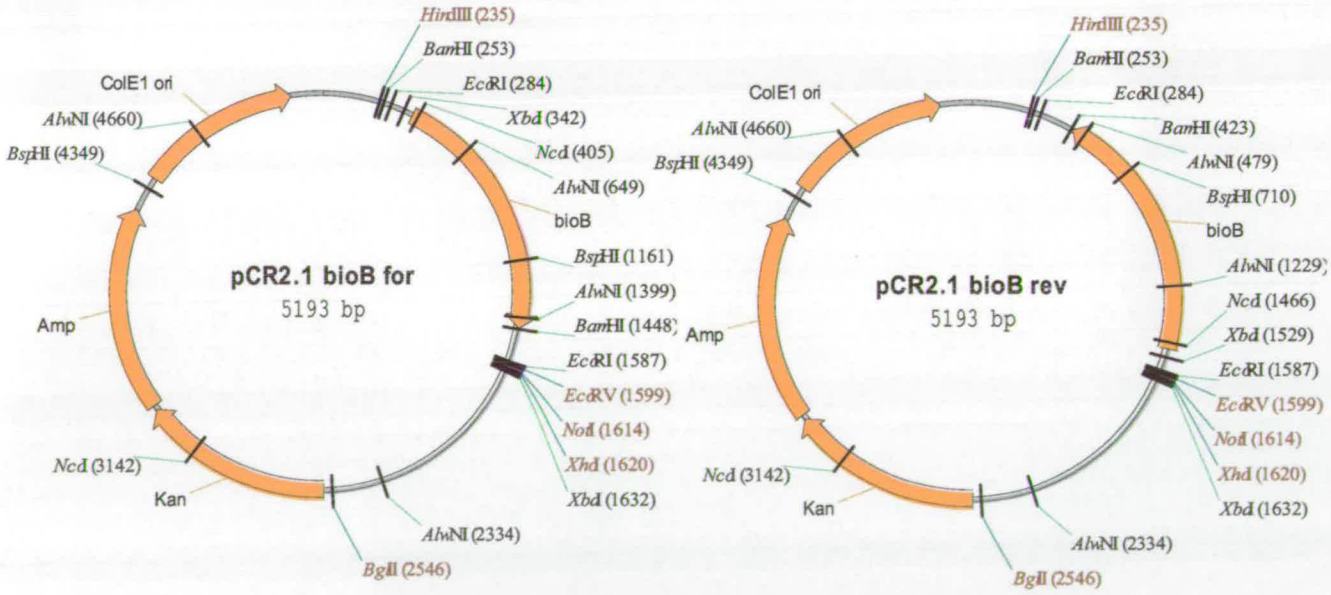


Figure 3.3 – Representation of the two possible products, forward and reverse configurations, of the PCR2.1 TA-TOPO cloning reaction to produce PCR2.1/bioB.

The residues in *E. coli* biotin synthase demonstrated to bind [Fe-S] clusters are cys53, cys57 and cys60 and this study complements the mutagenesis of these residues by mutatation of the gene to include multiple Cysteine to Serine mutations within the 'cys box' attempting to further disrupt the cluster binding capacity of the subsequent protein.

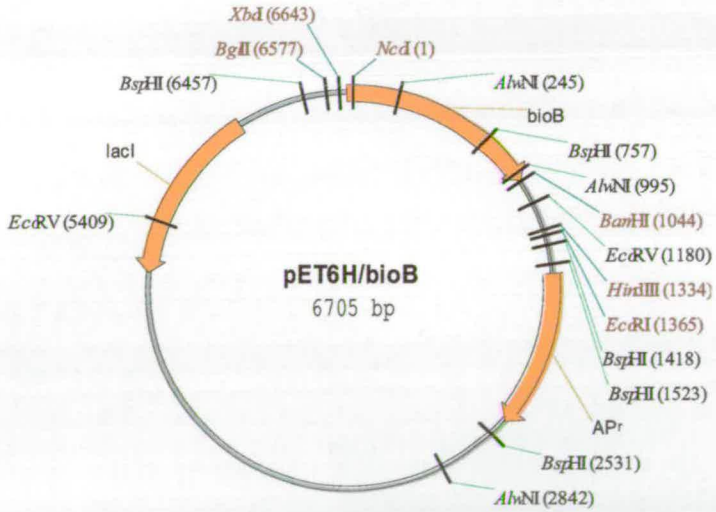
A complete complement of double and the triple mutants genes were constructed as shown in Figure 3.2 using the mega-primer method (165). The resultant PCR amplified mutant bioB genes were successfully ligated into the pCR2.1 TA-TOPO cloning vector (Invitrogen) as shown in Figure 3.3.

Other mutants constructed were lysine 49 mutations to glutamine and arginine for analysis of the cysteine desulfurase activity of biotin synthase as described in Section 3.3. These were constructed in the same manner as the cysteine mutants and are also shown in Figure 3.2.

The pCR2.1 vector containing the BS WT and mutant genes were subjected to restriction digest analysis with *Bam*HI to ascertain the direction of cloning and subsequently was successfully sequenced by automated DNA sequencing as described in Section 2.13 and the mutant sequence confirmed as those described in Figure 3.2, thus taken forward to expression and purification.

3.2 – Expression and Purification of 6HisBS WT and Mutants

The wild type and mutant genes as described in Figure 3.2 were excised using restriction enzymes from their corresponding pCR2.1/bioB plasmids and cloned in to pET6His, cut with the same enzyme, using the *Nco*I/*Bam*HI restriction site to produce pET6His/bioB and corresponding mutants as shown in Figure 3.4.The pET6His plasmid contains an ampicillin resistance gene (Ap^r) for selection and the *bio*B gene is under the control of the strong bacteriophage T7 transcription and translation signals and expression is induced with addition of IPTG. When cloning with 5'-terminus *Nco*I site the plasmid also yields a protein with an 8 residue extension at the N-terminus, namely MHHHHHHA, containing the 6 histidine tag which enables quick one step purification of protein via Immobilised Metal Affinity Chromatography (IMAC) on a Ni^{2+} resin..



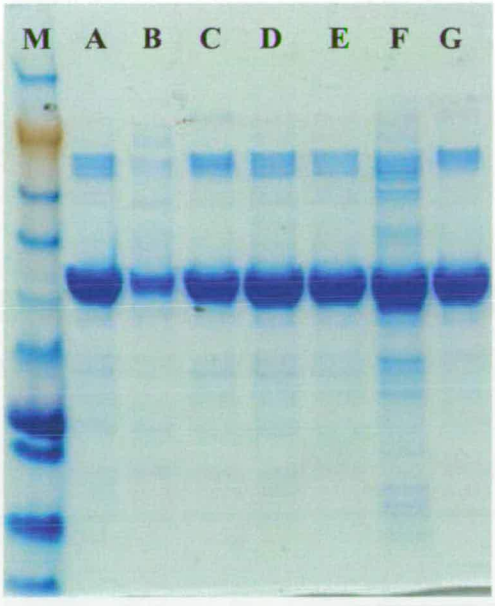
pET6His / bioB WT
pET6His / bioB C53/57S
pET6His / bioB C53/60S
pET6His / bioB C57/60S
pET6His / bioB C53/57/60S
pET6His / bioB K49Q
pET6His / bioB K49R

Figure 3.4 – Representation of expression plasmid used for biotin synthase and mutants. Contains an ampicillin resistance gene and T7 promotor controlling the protein expression and yields protein with N-terminal Histidine tag.

Production of wild-type and mutant proteins were performed in BL21(DE3) and HMS174(DE3) cells which both yeilded high expression levels. However, HMS174(DE3) cells were used routinely. Optimum expression levels were found using 2YT and ampicillin (100 μg/ml) broth at 37°C and when protein expression was induced using 1 mM IPTG as described in Section 2.17. Proteins were purified via IMAC as previously described (87) and eluted at approximately 200 mM imidazole.

The purity of the protein was judged by SDS-PAGE (Figure 3.5) and this showed that biotin synthase eluted greater than 95% pure in this single step and so for a majority of applications no further purification was performed. Routinely we were able to purify 30-40 mg of wild-type biotin synthase per litre of cell culture within a single day and the purified protein was stored in 10 mg aliquots at -20 °C.

Figure 3.5 – SDS-PAGE of elution fractions from purification of 6HisBS from HiTrap column. Fractyions 14 to 17 combined and dialysed to remove imidazole.. Lane M shows SeeBlue2 molecular weight marker (Invitrogen) [200kDa = Myosin, 116kDa = β-galactosidase, 97kDa = Phosphorylase b, 66kDa = BSA, 55kDa = Glutamic dehydrogenase, 36kDa = Lactate dehydrogenase, 31kDa = Carbonic anhydrase, 21kDa = Trypsin inhibitor, 14kDa = Lysosyme, 6kDa = Aprotinin, 3,5kDa = unresolved Insulin]



Label	Protein	Typical yields per litre culture (mg)
A	WT	20
B	C53/57S	15
C	C53/60S	20
D	C57/60S	10
E	C53/57/60S	10
F	K49Q	15
G	K49R	15

Figure 3.6 – SDS-PAGE of purified biotin synthase and mutants A to G and yields shown in table for 1 litre of culture. Lane M shows SeeBlue2 molecular weight marker (Invitrogen) [200kDa = Myosin, 116kDa = β-galactoside, 97kDa = Phosphorylase b, 66kDa = BSA, 55kDa = Glutamic dehydrogenase, 36kDa = Lactate dehydrogenase, 31kDa = Carbonic anhydrase, 21kDa = Trypsin inhibitor, 14kDa = Lysosyme, 6kDa = Aprotinin, 3,5kDa = unresolved Insulin]

All biotin synthase mutants were purified using the same method and elution profiles were comparable to wild-type biotin synthase however total protein yield generally decreased between wild-type and mutant expression and varied from 15-30 mg per litre of cell culture depending on mutant (inset Figure 3.6). All proteins were again judged for purity by SDS-PAGE (See Figure 3.6) and concentrations estimated by Bradford assay (166) using BSA as a standard.

After purification all proteins were dialysed against Buffer C (50 mM Tris, pH7.5) at 4 °C over night. Mutant proteins BS C57/60S and BS C53/57/60S became unstable when NaCl and imidazole concentrations were reduced and completely precipitated within 2 hours of dialysis so these mutants were not subjected to further analysis. We believe these mutations may affect the tertiary or quaternary structure of the protein causing protein instability through aberrant, misfolding and/or denaturation followed by precipitation.

3.3 – UV-Vis characteristics of biotin synthase and mutants

After purification and dialysis against 50 mM Tris-pH7.5, UV-Vis absorbance spectra were measured for wild-type biotin synthase and mutants C53/57S, C53/60S, K49Q and K49R. The scan was performed using the same concentrations of the proteins between 250 nm and 600 nm and the spectrometer blanked with a portion of the used dialysis buffer. Figure 3.7 shows the UV-Vis spectra for each soluble mutant and also shows an overlay of all spectra which was corrected to 280 nm. This clearly enables visual comparison of the protein UV-Vis absorbance characteristics of the purified biotin synthase proteins. The observed absorbance between 420-500 nm is responsible for the observed deep red colour of biotin synthase. The spectral characteristics associated with aerobically purified 6HisBS are observed in all mutants however they are significantly reduced in the C53/57S mutant. The spectra show absorbance maxima at 280 nm, 330 nm, 460 nm a distinct plateau leading from 400nm to 450nm. The maxima have been compared as shown in Figure 3.7 highlighting ratios of $^{330nm}/_{280nm}$, $^{460nm}/_{280nm}$ and $^{460nm}/_{330nm}$. The comparison of the absorbance at 330nm and 280nm shows a marked decrease in the 330nm maxima for the BS C53/57S and a smaller decrease for BS C53/60S whereas little difference is observed between wild-type and K49 mutants. As the C53,57 and 60 have been shown to be involved in the binding of the redox active [4Fe-4S] cluster (87, 86) we propose that this spectral feature is primarily derived from the formation of this or an intermediate cluster, for example the [2Fe-2S] present when biotin synthase is prepared aerobically. In contrast the 460 nm to 280 nm ratio shows little variation throughout the wild-type and mutants. The ratios of 460 nm to 330 nm seem to be maintain constant between all the mutants with exception again being the C53/57S mutant which shows an increase in this absorbance ratio of the 460 nm and 330 nm maxima. In conclusion these mutations have had little effect on the UV-Vis characteristics of the purifed protein excepting the C53S/C60S mutant. Thus we must conclude that the mutations made to the protein, C53S/C57S, K49Q and K49R have not greatly effected the [Fe-S] cluster characteristics compared to WT in this aerobic environment. In contrast the C53S/C60S mutant shows deviation from this

characteristic UV-Vis profile suggesting that the binding of the aerobic [2Fe-2S] cluster is greatly hampered by this double mutation.

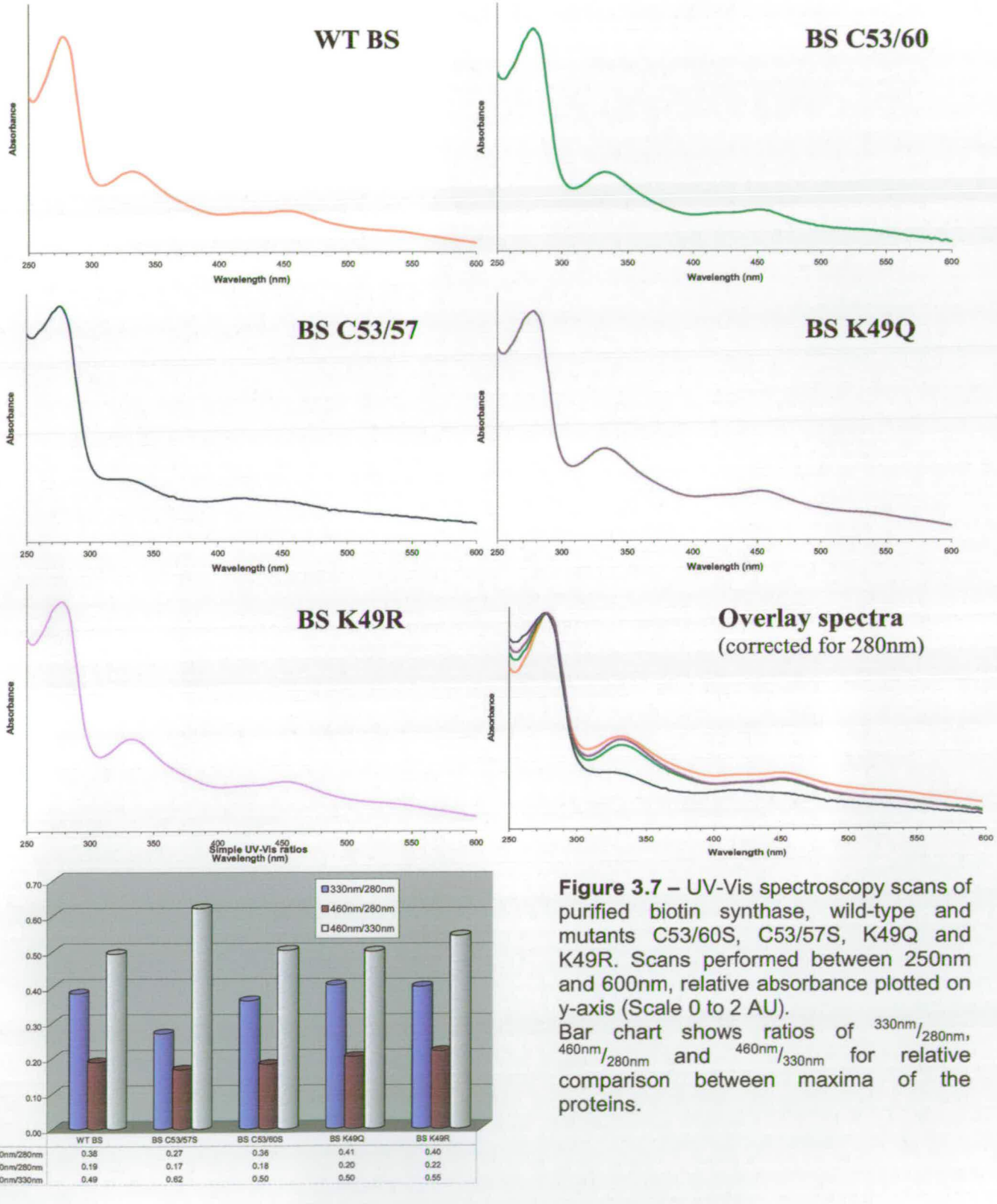


Figure 3.7 – UV-Vis spectroscopy scans of purified biotin synthase, wild-type and mutants C53/60S, C53/57S, K49Q and K49R. Scans performed between 250nm and 600nm, relative absorbance plotted on y-axis (Scale 0 to 2 AU).
Bar chart shows ratios of $^{330nm}/_{280nm}$, $^{460nm}/_{280nm}$ and $^{460nm}/_{330nm}$ for relative comparison between maxima of the proteins.

3.4 – Mass Spectroscopy of Biotin Synthase and Mutants

Purified proteins were all subjected to LC-MS analysis to confirm their theoretical mass as calculated using VectorNti software. Proteins were loaded (30 µl) from an approximately 1 mg/ml solution and run using a standard LC-MS run using a 20 min reverse phase gradient optimised using Myoglobin, as described in Section 2.28. The resultant ion-envelope spectra were deconvoluted by MassLynx algorithms Transform and MaxEnt, the results spectra for the MaxEnt deconvolution are shown in Figure 3.9. The table imbedded in Figure 3.9 shows the obtained average molecular weight for all proteins using both deconvolution algorithms.

The major mass species of all proteins were within 0.01 % of the calculated theoretical masses from VectorNti. Un-resolvable higher mass species were observed as high mass tailing up to approximately 500 Da above the calculated mass. The high mass tail observed in all spectra can be attributed to specific and non specific protein adducts maintained during the LC-MS process. Non specific adducts such as sodium and potassium could account for some of this poorly defined peak shape. However, the ion-envelope topology is intriguing and some closer analysis reveals some interesting qualitative information.

Over the last 10 years the analysis of entire proteins by mass spectrometry has developed from a 'molecular weight conformation tool' into one where qualitative structural information can be elucidated from the over topology of observed ion envelopes. Several important studies have been conducted looking at overall protein folds and how they relate to observed ion-envelopes (167, 168, 169, 170, 171). In many of these studies well characterised proteins have been analysed varying buffers incubation time etc. These studies show that the more structured the protein chain, the less projected area is available for accommodation of charge (11). Experimentally, a protein with 3-dimensional structure (compared to the same protein in a random coil) display an ion envelope biased toward the high m/z region due to the few charges being maintained on the protein surface. This is shown in Figure 3.8.

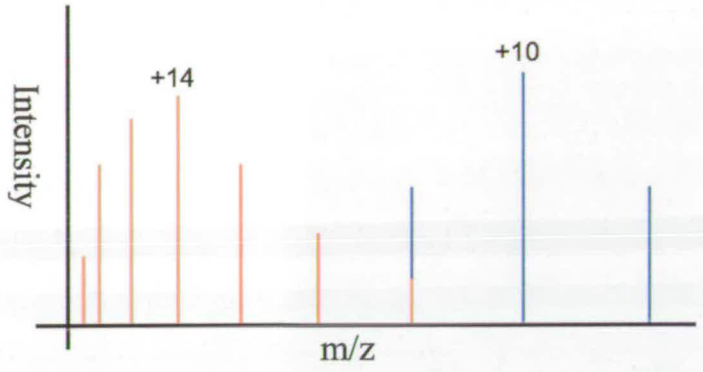
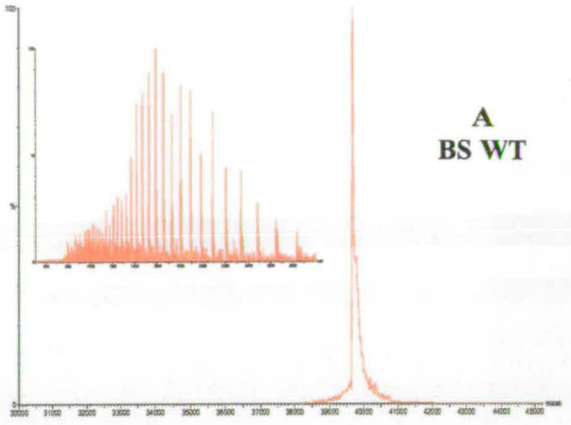


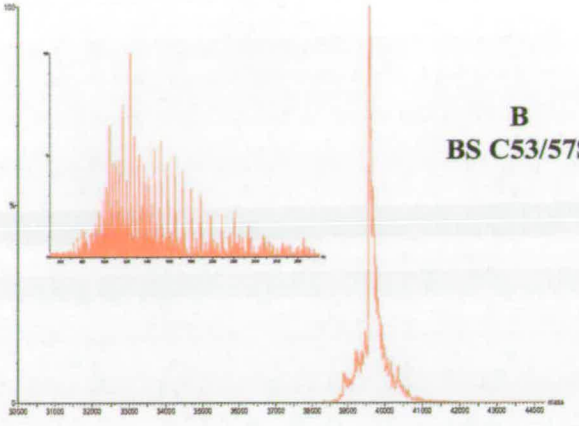
Figure 3.8 – Diagram showing general topology of theoretical ion-envelope (red) of a random coil protein compared to the same protein with 3-dimensional structure maintained (blue). The less random coil like the protein chain the more likely it is to observe the lower charge states at higher m/z values. Based on Kaltashov et. al. (11)

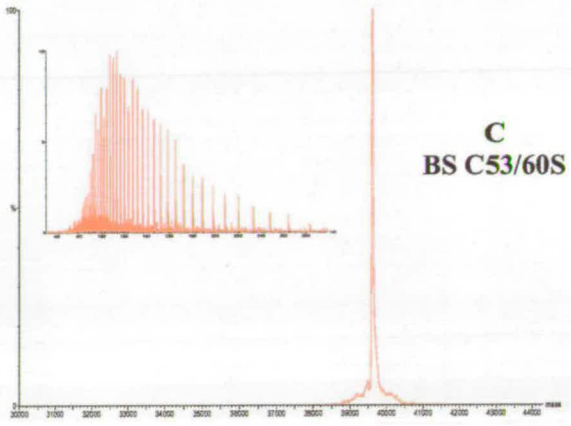
A simple visual comparison of the acquired ion-envelopes of BS and its mutants shows that several differences are observed between them. Both double cysteine mutants C53/57S and C53/60S (B and C respectively in Figure 3.9) show a predominance for the lower m/z range, maximum at approximately 1000 m/z which is a charge state in the region of +40. In contrast, the BS wild type and BS_K49R (A and D respectively in Figure 3.9) proteins show envelope maxima in the region of 1600 m/z which would equate to charge states of approximately +25. Clearly, even though the protein chain as not altered considerably between these species their behaviour during ionisation and MS analysis differs greatly. If we apply the ideas discussed above we can conclude that in comparison to the double cysteine mutants the wild type and K49R proteins have more structural integrity during the LC-MS experiment. The observed masses for all the proteins show no evidence for major modifications on the majority of protein analysed. However, we can easily conclude that the conserved cysteine residues in BS (C53, C57 and C60) may play an important structural role whether a cluster is bound to them or not. This is particularly interesting as the crystal structure seems to show that they are present on a loop region and the authors suggest that structure in this region is maintained by the cluster (16). The observed ion-envelope for the last species analysed, K49Q, seems

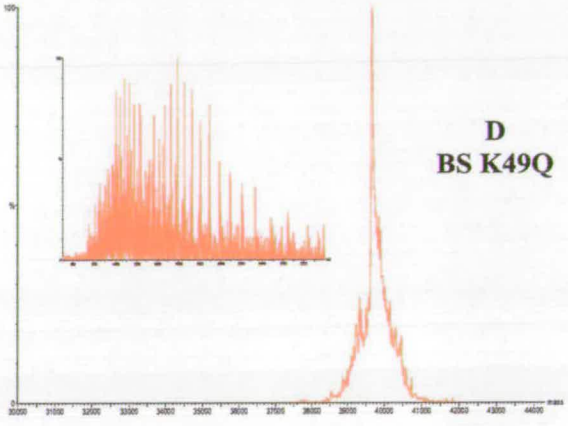
to show a combination of the topologies described abovem where it has 2 maxima regions, one around 1000 m/z and one around the 1600 m/z region. We must conclude that this preparation contains at least 2 major structures, one which is like the WT and one which displays the more random coil like structure which may interchange in solution.

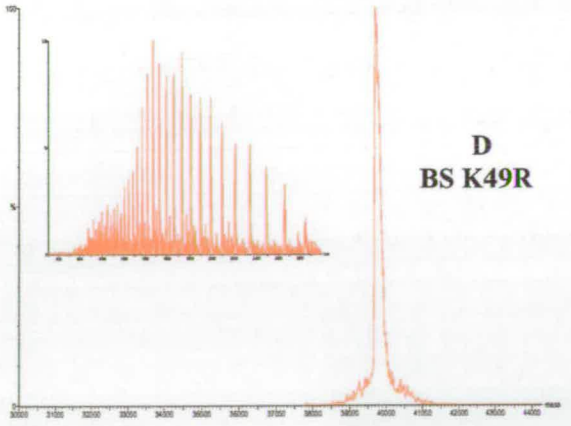
Although it is impossible to conclude real structural/conformational information from these simple observations it is valuable to compare how these simple amino acid changes within the BS sequence at sites recognised as important can affect the gas phase structure in such a obvious manner.











Spectra	Theoretical	Transform	Error	MaxEnt	Error
A	39,670.63	39,673.00	+2.37	39,671.00	+1.37
B	39,638.49	39,641.00	+2.51	39,640.00	+1.51
C	39,638.49	39,642.00	+3.51	39,639.00	+0.51
D	39,670.63	39,664.00	-6.63	39,670.00	-0.63
E	39,698.54	39,704.00	+5.36	39,703.00	+4.46

Figure 3.9 – LC-MS spectra of wild-type and mutant biotin synthase purified proteins. Traces shown are deconvoluted from ion-envelope (inset) using the Transform algorithm. Theoretical and calculated masses, Transform and MaxEnt algorithms, and Da errors are shown in table.

3.5 – Identification of the Pyridoxal-Phosphate Binding Residue of *E. coli* Biotin Synthase

3.5.1 – Introduction

Prior to publication Prof. Marc Fontecave visited our laboratory in 2001 and presented data that suggested BS had PLP-dependant cysteine desulferase activity. His data clearly shows that their biotin synthase preparations did convert free cysteine to alanine and sulphide in the presence of PLP albeit at a low rate. After a discussion on the possible experiments we could do to further investigate the hypothesis he agreed to supply some protein from his laboratory.

The initial aims of this study were, firstly, to unambiguously identify any covalent link between BS and PLP and secondly, to explore if the conserved Lys49 residue of biotin synthase is involved in PLP binding. We intended to utilise a variety of mass spectrometry instrumentation available for this purpose. We also wanted to exploit our experience with PLP enzymes gained during analysis of other members of the biotin biosynthetic pathway (AONS and DANS) to study the UV-Vis characteristics of the PLP modified biotin synthase species in the presence of cysteine.

3.5.2 – PAWS digestion analysis

The theoretical sequence of biotin synthase was modified to contain the histidine tag at the N-terminus (MHHHHHHA) and this was used for theoretical enzymatic digestion with endoproteinases. Trypsin and Asp-N enzymes were chosen to facilitate the further study of the PLP modification of biotin synthase. This analysis produced a list of M^{+1} ions corresponding to enzymatically derived peptides from biotin synthase (156, 154). Experimentally only specific peptides are observable due to m/z limits of the analysis. Specifically peptides of m/z less than 500 would be hard to observe in the MALDI-ToF spectra as they are likely to be indistinguishable in this highly populated region of the spectra a major constituent of which will be the abundant matrix ions observable in the 50-500 m/z region. Similarly ions of m/z greater that 3000 would be difficult to observe in the usual peptide mass fingerprint (PMF) spectra as ionisation tends to requires higher laser energies. Use of high laser energies tends to cause peptide fragmentation and as such

produces additional ambiguous peptide species in the fingerprint spectrum. During the nanoESI-Q-Tof analysis ions with m/z less than 500 were found to be hard to observe due to formation of high intensity H_2O clusters, chemical noise and abundant small molecular contaminants. Although nanoESI-Q-Tof would tend to allow observation of larger species due to multiple charging of peptides, observation of $M^{>+1}$ species, definitive identification was impossible for $M^{>+3}$ due to a resolution limit and as such distinguishing +4 and +5 charge states from each other was unreliable.

Shown in figure 3.10a and 3.10b are tables showing all theoretical peptides derived from digestion of *E. coli* biotin synthase with the endoproteinase trypsin and AspN respectively and their related m/z value for different charge states.

3.5.3 – In Gel Digestion of Biotin Synthase and Preparation for Mass Spectrometry

Two sources of biotin synthase were used at this stage our own 6His-BS and a wild-type BS obtained from Prof. Marc Fontecave which had been purified and the iron-sulfur cluster(s) reconstituted anaerobically in the presence of $Fe(NH_4)_2(SO_4)_2$ and Na_2S in 6 molar excess (in this thesis termed BS_MF). This procedure has been previously described (89, 86, 109). The samples were then anaerobically incubated for 4 hours with 5 molar excess of PLP and excess PLP removed by 2 x 1 litre dialysis against 100 mM Tris-HCl pH 7.5 and finally treated with 1 mM $NaBH_4$ for 5 min at 25 °C. (92, 91). Aliquots were analysed by electrophoresis using Nu-PAGE gel (Invitrogen) as described in Section 2.15 and Mark12 molecular weight marker used to estimate protein MW (see Figure 3.11). Bands were excised and subjected to in-gel enzymatic digestion as described in Section 2.24.1. The resultant protein digest was dried to a pellet of peptides and salts visible in eppendorf tube. For all analysis samples were reconstituted in 20 µl of H_2O + 0.1 % TFA and desalting carried out using C18 Reverse Phase ZipTips (Millipore) as manufacturers protocol and peptides eluted using 5µl of 50% acetonitrile, 0.1% TFA in ddH_2O (154).

Peptide No..	m/z (M+1)	m/z (M+2)	m/z (M+3)	m/z (M+4)	m/z (M+5)	Sequence
1	2554.030	1277.515	852.010	639.258	511.606	MAHRPRWTLSQVTELFEKPLL
2	**1768.970**	884.985	590.323	442.993	354.594	DLLFEAQQVHRQHF
3	**2200.510**	1100.755	734.170	550.878	440.902	DPRQVQVSTLLSIKTGACPE
4	6192.030	3096.515	2064.677	1548.758	1239.206	DCKYCPQSSRYKTGLEEARLMEVEQVLESARKAKAAGSTR FCMAAWKNPHER
5	4214.900	2107.950	1405.633	1054.475	843.78	DMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGL
6	**938.970**	469.985	313.657	235.493	188.594	DYYNHNL
7	**2405.620**	1203.310	802.540	602.155	481.924	DTSPEFYGNIITTRTYQERL
8	**860.980**	430.990	287.660	215.995	172.996	DTLEKVR
9	**1861.160**	931.080	621.053	466.040	373.032	DAGIKVCSGGIVGLGETVK
10	**3468.180**	1734.590	1156.727	867.795	694.436	DRAGLLLQLANLPTPPESVPINMLVKVKGTPLA
11	248.220	124.610	83.407	62.805	50.444	DN
12	134.110	67.555	45.370	34.278	27.622	D
13	233.240	117.120	78.413	59.060	47.448	DV
14	352.370	176.685	118.123	88.843	71.274	DAF
15	6520.560	3260.780	2174.187	1630.890	1304.912	DFLIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMCFMAGANSIF YGCKLLTTPNPEE
16	**262.290***	131.645	88.097	66.323	53.258	DK
17	**2183.550**	1092.275	728.517	546.638	437.51	DLQLFRKLGLNPQQTAVLAG
18	**1801.970**	901.485	601.323	451.243	361.194	DNEQQQRLEQALMTP
19	**235.220****	118.110	79.073	59.555	47.844	DT
20	**1030.080**	515.540	344.027	258.270	206.816	DEYYNAAAL

Figure 3.10a – Table of theoretical peptides derived from a complete digestion of *E. coli* 6HisBS by endoproteinase Asp-N. Compiled using PAWS software and sequence manipulated using VectorNti. Masses observed during MALDI analysis in **Bold** and peptides observed during QTOF analysis underlined. Non observed theoretical peptide masses in grey. * observed in combination with peptide 17 as miss cleavage. ** observed in combination with peptide 18 as miss cleavage.

Peptide No..	m/z (M+1)	m/z (M+2)	m/z (M+3)	Sequence
1	767.930	384.465	256.643	MAHRPR
2	3142.620	1571.810	1048.207	WTLSQVTELFEKPLLDLLFEAQQVHR
3	**799.860**	400.430	267.287	QHFDPR
4	1216.460	608.730	406.153	QVQVSTLLSIK
5	924.040	462.520	308.680	TGACPEDCK
6	**840.930**	420.965	280.977	YCPQSSR
7	**310.370**	155.685	104.123	YK
8	**775.840**	388.420	259.280	TGLEAER
9	**1404.620**	702.810	468.873	LMEVEQVLESAR
10	147.200	74.100	49.733	K
11	218.280	109.640	73.427	AK
12	**562.600**	281.800	188.200	AAGSTR
13	**914.130**	457.565	305.377	FCMGAAWK
14	**652.690**	326.845	218.230	NPHER
15	**1538.820**	769.910	513.607	DMPYLEQMVQGVK
16	**2098.420**	1049.710	700.140	AMGLEACMTLGTLSESQAQR
17	**3074.330**	1537.665	1025.443	LANAGLDYYNHNLDTSPEFYGNIITTR
18	**696.740**	348.870	232.913	TYQER
19	**718.820**	359.910	240.273	LDTLEK
20	274.340	137.670	92.113	VR
21	503.570	252.285	168.523	DAGIK
22	1319.560	660.280	440.520	VCSGGIVGLGETVK
23	290.300	145.650	97.433	DR
24	2530.080	1265.540	844.027	AGLLLQLANLPTPPESVPINMLVK
25	**246.330**	123.665	82.777	VK
26	**1881.990**	941.495	627.997	GTPLADNDDVDAFDFIR
27	**630.760**	315.880	210.920	TIAVAR
28	1098.370	549.685	366.790	IMMPTSYVR
29	503.580	252.290	168.527	LSAGR
30	2704.110	1352.555	902.037	EQMNEQTQAMCFMAGANSIFYGCK
31	**1257.380**	629.190	419.793	LLTTPNPEEDK
32	**791.920**	396.460	264.640*	DLQLFR
33	**147.200**	74.100	49.733	K
34	**2181.370**	1091.185	727.790	LGLNPQQTAVLAGDNEQQQR
35	2130.330	1065.665	710.777	LEQALMTPDTDEYYNAAAL

Figure 3.10b – Table of theoretical peptides derived from a complete digestion of *E. coli* 6HisBS by endoproteinase trypsin. Compiled using PAWS software and sequence manipulated using VectorNti.Masses observed during MALDI analysis in **Bold** and peptides observed during QTOF analysis underlined. Non observed theoretical peptide masses in grey. * observed in combination with peptide 32 as miss cleavage.

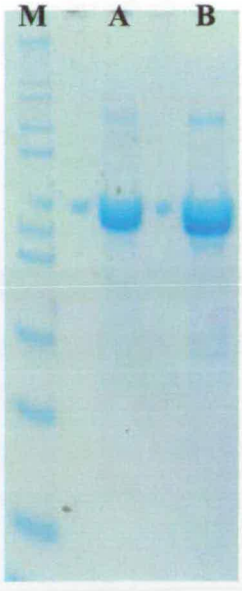


Figure 3.11 – Nu-PAGE gel showing purified biotin synthase bands excised for subsequent in-gel digestion and Mass Spectrometry analysis, bands halfed and one half digested with Trypsin and other with AspN. **A** – 6His_BS, **B** – BS_MF. **M** – molecular weight marker (Mark12, Invitrogen) [200kDa = Myosin, 116kDa = β–galactoside, 97kDa = Phosphorylase b, 66kDa = BSA, 55kDa = Glutamic dehydrogenase, 36kDa = Lactate dehydrogenase, 31kDa = Carbonic anhydrase, 21kDa = Trypsin inhibitor, 14kDa = Lysosyme, 6kDa = Aprotinin, 3,5kDa = unresolved Insulin]

3.5.4 – MALDI MS Peptide Fingerprint of Biotin Synthase

Desalted enzymatically derived peptide mixures of biotin synthase were analysed with the Applied Biosystems DE^{TM}STR MALDI-ToF as described in 2.29 to obtain a fingerprint spectra for direct comparison with the theoretical PAWS derived peptides. The in-gel digestion process should theoretically yield a complete complement of peptides, fully reduced with no intra-molecular disulfides. It is in practice very rare to observe all theoretical peptides due to experimental factors described above as well as multiple miss cleavages instrument limits of detection and non ionisable peptide chains.

Trypsin Digest

The detected fingerprint spectra of an in-gel trypsin digestion of biotin synthase is shown in Figure 3.13. Peaks are labelled with the chronological peptide number N-terminus to C-terminus as described. The total residue coverage was 207 residues of 346 total (>60%), this corresponding to 22 of the 35 theoretical cleavage peptides

(>65%) which demonstrated extremely high coverage of the entire protein sequence from N to C terminus (See Figure 3.12). Six of the 13 peptides not observed in the spectra were not practically observable due to their mass being ≤500 Da. During analysis a 500 m/z gate was set with the instrument to increase sensitivity of species in the range 500 to 3500 m/z, the area which tends to provide the most significant and specific data. The remaining 7 theoretical peptides not observed have predicted m/z values within the range 500 to 3500 m/z. An explanation could be due to post-translational modifications, which will alter the mass of the peptide and thus shift the m/z resulting in a non recognisable peptide species. Secondly a certain number may not ionise easily using MALDI or they may require a different matrix.

In addition 4 species were observed in this analysis which originate from a single miss cleavage of a lysine (K) or arginine (R) residues. These miss-cleaved species observed at m/z 1066.1002 corresponding to peptide 7 and 8 and deriving from a miss cleavage at lysine67 and a species at m/z 2108.050 is derived from a miss cleavage at residue lysine228 and combines peptides 25 and 26. The species observed at 2029.703 m/z contains peptides 31 and 32 and is similarly derived from a miss cleavage, in this case at lysine300. The final miss-cleaved species observed at 2308.923 m/z derives from a miss cleavage at lysine307 and combines peptides 33 and 34.

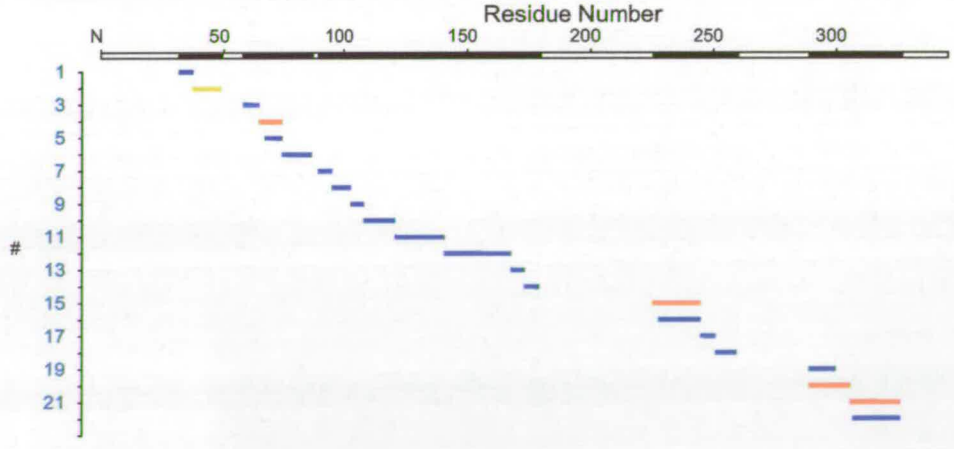


Figure 3.12 – Coverage map of peptides observed during in-gel trypsin digestion MALDI-ToF of biotin synthase. Miss cleavages shown in Red. K49 containing peptide shown in yellow

As shown in Figure 3.13 two species are observed that corresponds to peptide 4 (Tryp4-WT) and peptide 4 plus the addition of 231 Da (Tryp4-WT+PLP) which is consistent with the hypothesis of Ollagnier-de-Choudens *et al.* (91, 92) of PLP modification of K49. This observation strongly supports the modification of this highly conserved lysine residue (lysine49 in *E. coli* biotin synthase) by PLP implicated in the proposed cysteine desulfurase activity reported by Ollagnier-de-Choudens *et al.*

Tryp4-WT -	QVQVSTLLSIK	M+1 = 1216.460 m/z
Tryp4+PLP-	QVQVSTLLSIK-PLP	M+1 = 1445.489 m/z

The low relative intensity of the species at 1216 m/z and 1445 m/z (Tryp4-WT and Tryp4-WT+PLP respectively) compared to other biotin synthase derived peptides during this analysis was seen as a minor issue and as such a further study using MS/MS techniques was used to confirm this initial identification of the PLP modified residue as lysine 49.

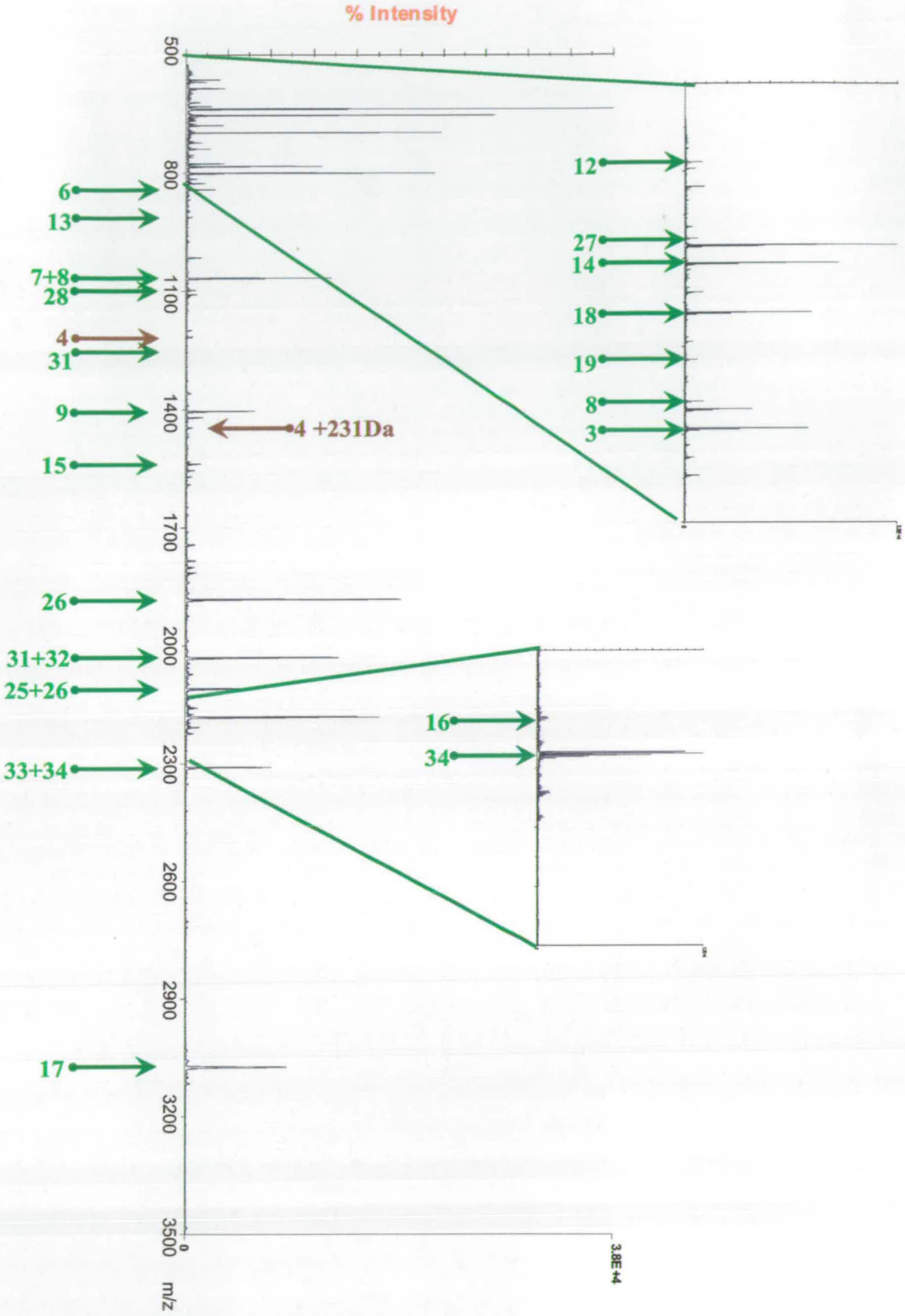


Figure 3.13 – MALDI-ToF spectra of in-gel trypsin digested biotin synthase , Peptides matched with theoretical (See Figure 3.10) shown in **green**, peptide containing lysine 49 and corresponding +231Da species shown in **burgundy** .

AspN Digest

The peptide fingerprint spectra shown in Figure 3.15 shows all observed peptides N to C-terminus when wild-type biotin synthase was in-gel digested with endoproteinase AspN. The total residue coverage observed was 195 of 346 residues (>56%) which correlates to 12 out of 22 (>54%) theoretical peptides derived from a biotin synthase digestion with AspN. The observed peptides show fairly high linear coverage of the protein sequence as shown in Figure 3.14.

The much of the unobserved portions of the protein sequence was due to the small peptides with m/z of less than 500 therefore lie outside the observable region of the MALDI-TOF analysis. However 2 large peptides are unobserved and correspond to the major uncovered regions in the coverage map (Figure 3.14, shown in grey).

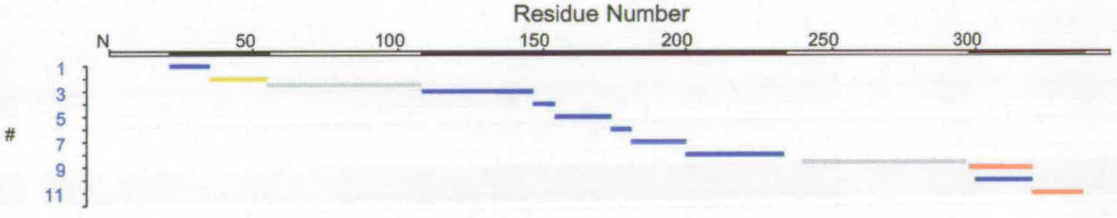


Figure 3.14 – Coverage map of peptides observed during in-gel AspN digestion MALDI-ToF of biotin synthase. Miss cleavages shown in Red. K49 containing peptide shown in yellow. Peptides (>6kDa) observed using Sinapinic acid matrix in separate spectra shown in grey.

The digest mixture was analysed using sinapinic acid as the Matrix and the laser intensity increased correspondingly. This analysis allowed observation of these two >50 residue peptides however many of the other peptides in the mixture were not observed in this analysis (data not shown). Two relatively low intensity species were observed corresponding to single miss cleavages. The species at 2443 m/z derives from a miss cleavage at Aspartic acid 301 and constitutes peptide 16 and 17 and the species at 2015 m/z is derived from a miss cleavage at Aspartic acid 335 and contains peptides 18 and 19.

In concurrence with the trypsin digested biotin synthase MALDI-ToF fingerprint we observed a low intensity species that corresponds to the peptide containing the highly conserved lysine49 (AspN3-WT), as well as a species corresponding to this peptide with the addition of 231Da (AspN3-WT+PLP) as shown in Figure 3.15. This observation is again consistent with the hypothesis that this Lys49 residue is modified by PLP as shown.

AspN3-WT -	DPRQVQVSTLLSIKTGACPE	M+1 = 2200.510 m/z
AspN3-WT+PLP-	DPRQVQVSTLLSIKTGACPE -PLP	M+1 = 2431.539 m/z

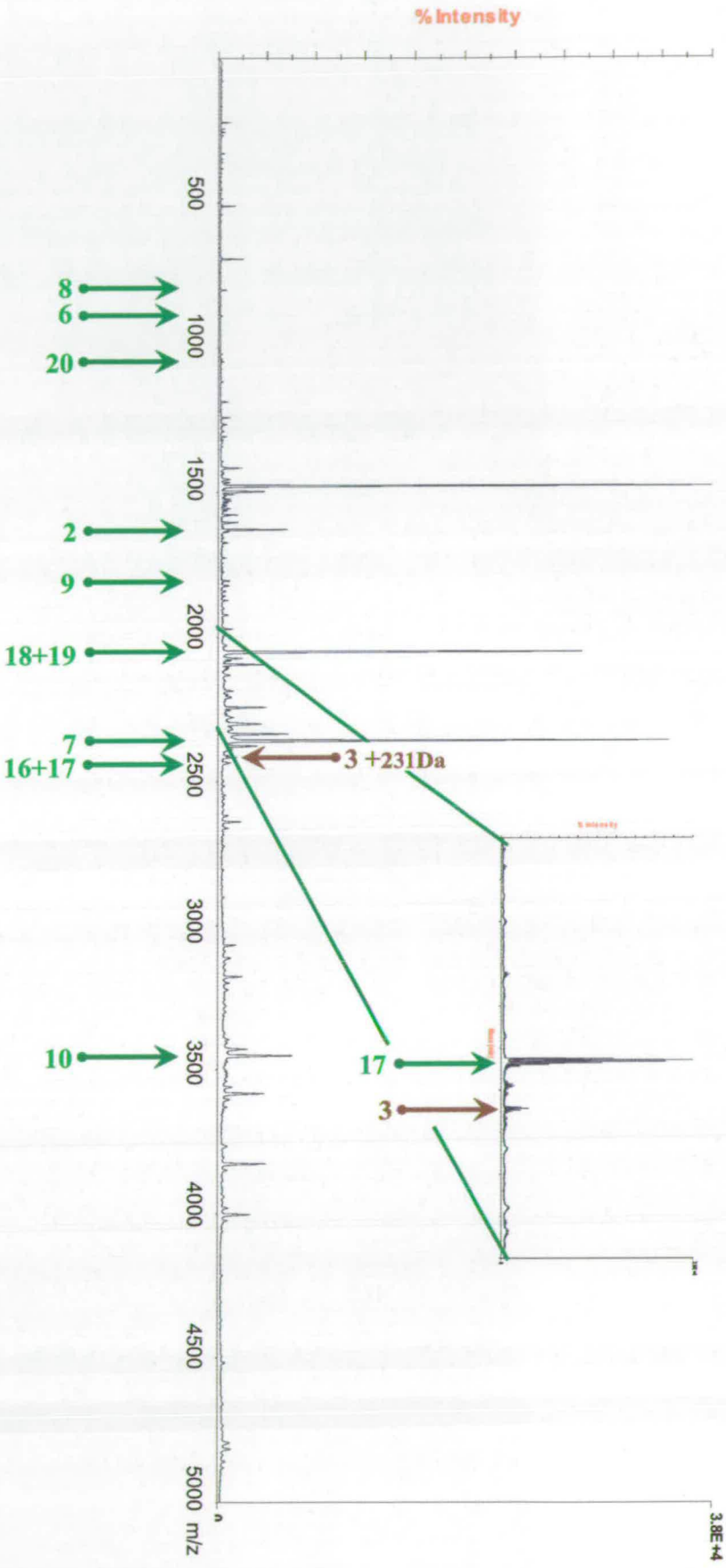


Figure 3.15 – MALDI-ToF spectra of in-gel AspN digested biotin synthase , Peptides matched with theoretical (See Figure 3.10) shown in **green**, peptide containing lysine 49 and corresponding +231Da species shown in **burgundy** .

3.5.5 – Q-ToF Peptide Fingerprint of Biotin Synthase

To complement the analysis of the protease fragments with MALDI we performed analysis of the peptides using a Q-Tof instrument. We hoped that this further analysis would generate a range of multiply charged species that would complement the MALDI analysis as well as providing enhanced sequence coverage. This analysis would also enable optimisation of the instrument for the further MS/MS analysis shown in section 3.57.

The samples used for Q-ToF mass spectrometry were prepared from the same gel-digestion as those used for MALDI-ToF analysis. Due to intolerance of salts in ESI ionisation, extra sample washing during the micro reverse phase desalting step was performed (C18, ZipTip (Millipore) washed with 6 column volumes).

The mass spectra observed of trypsin-digested biotin synthase using the Q-ToF in ToF-MS mode is shown in Figure 3.19a. All observed peptides are annotated and labels are used to describe the charge state at which they were observed correlating to the trypsin fragments in Figure 3.10b. The total amino acid coverage observed was 230 of 346 residues (>66%) which correlates to 16 out of 35 (>45%) theoretical peptides derived from a biotin synthase digestion with trypsin. The observed peptides show good linear coverage of the protein sequence (See Figure 3.16) and the mass of the unobserved peptides tends toward the lower masses (100 to 1000 Da). Due to instrument parameters used no species with m/z less than 500 were observed, as most trypsin peptides are doubly or more charged, we unlikely to observe fragments below 1000 Da. Seven peptides were observed as singly charged, three of which could not be observed as multiply charged species.

Importantly for the confirmation of the site of PLP modification, the peptide containing the theoretical binding lysine residue, Tryp4-WT containing K49, is clearly visible as a doubly charged species at 608.73 m/z and a species is also observed corresponding to Tryp4-WT+PLP is present at 724.23 Da as shown in Figure 3.19a.

Tryp4-WT	QVQVSTLLSIK	608.730 m/z (M^{+2})
Tryp4-WT+PLP	QVQVSTLLSIK-[PLP]	724.029 m/z (M^{+2})

The observation of these species is further evidence that lysine 49 is modified by PLP and further corroborates observations made during the MALDI analysis of the trypsin digest. Multiply charged peptides typically allow observation of both halves of the peptide chain after a single bond cleavage and as such the +2 charge state makes these species ideal for Collision Induced Dissociation (CID) MS/MS sequencing. This should allow further confirmation of the identity of this modified species and therefore verify the site of PLP modification as Lysine 49. The Q-Tof MS/MS data is shown in Section 3.5.7.

Q-ToF-MS analysis was also performed using the Asp-N digestion and the result of this is shown in the spectra shown in Figure 3.19b. The MS spectra observed again highlights observed peptides and their apparent charge state based on C^{13} isotope distribution. The total residue coverage observed was 181 of 346 residues (>52%) which equates to 9 out of 20 (45%) theoretical peptides derived from a biotin synthase digestion with AspN. The observed peptides show acceptable linear coverage of the protein sequence (See Figure 3.18).

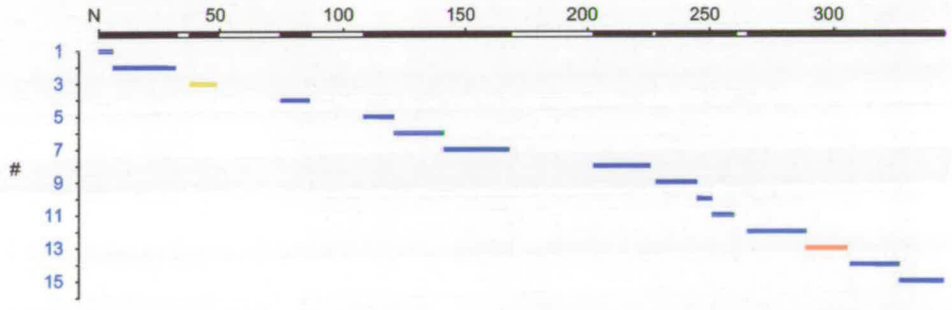


Figure 3.16 – Coverage map of peptides observed (blue) during in-gel trypsin digestion Q-ToF MS of biotin synthase. Miss cleavages shown in Red. K49 containing peptide shown in yellow

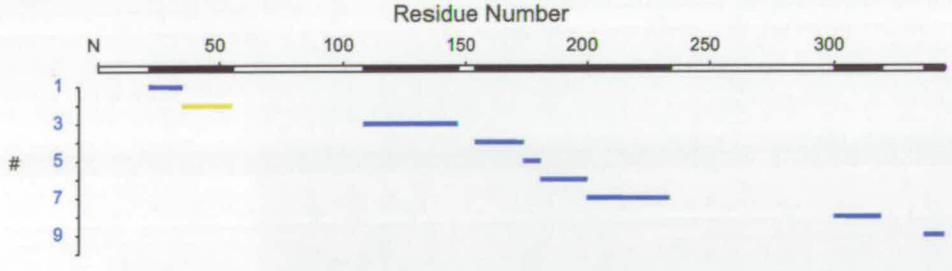


Figure 3.18 – Coverage map of peptides observed during in-gel AspN digestion Q-ToF MS of biotin synthase. K49 containing peptide shown in yellow

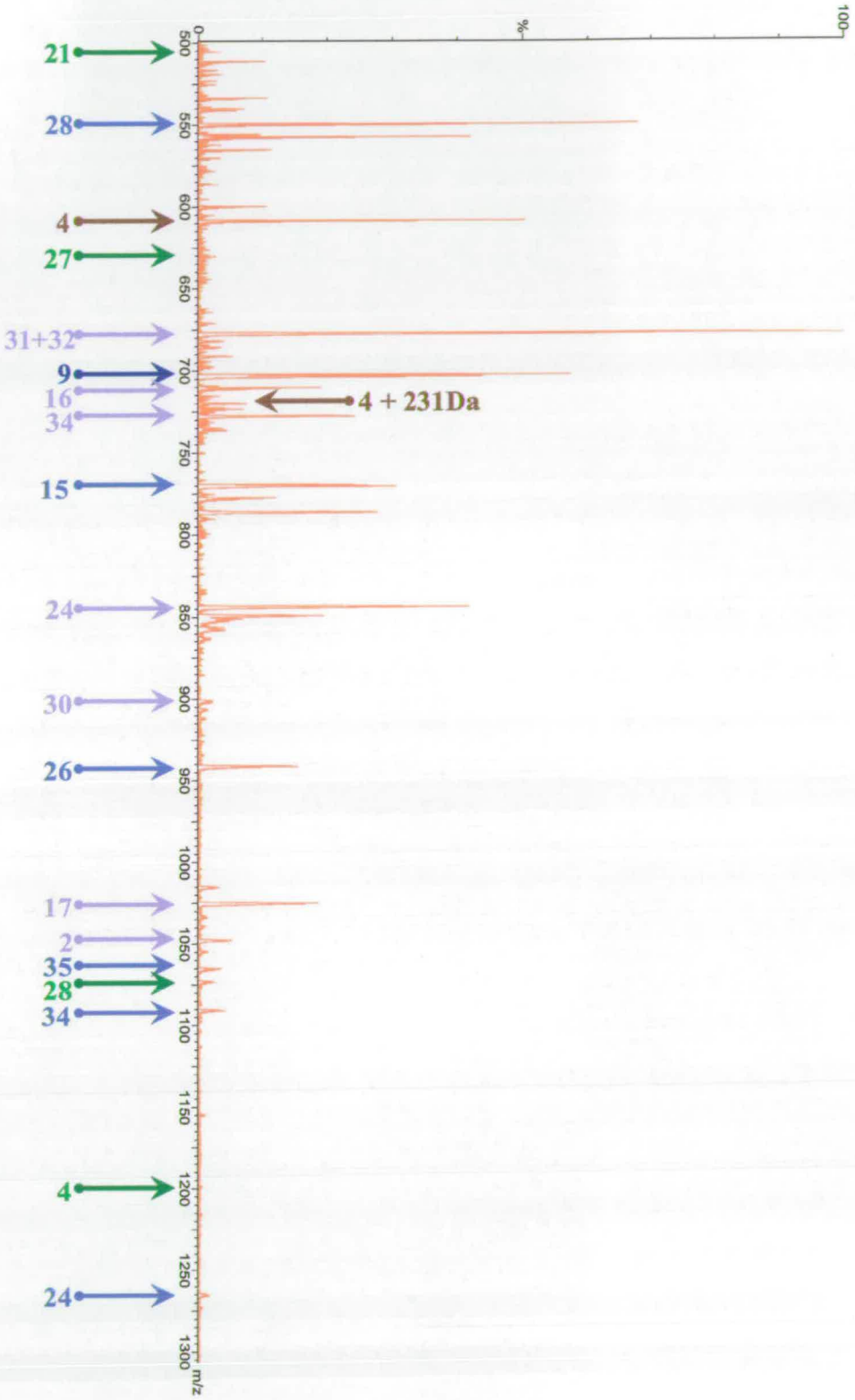


Figure 3.19a – Q-ToF MS spectra of in-gel trypsin digested biotin synthase , Peptides matched with theoretical (See Figure 3.10) labelled and colour coded as to charge state as follows, M^{+1} in **green**, M^{+2} in **blue**, M^{+3} in mauve and peptide containing lysine 49 and corresponding +231Da species shown in **burgundy** .

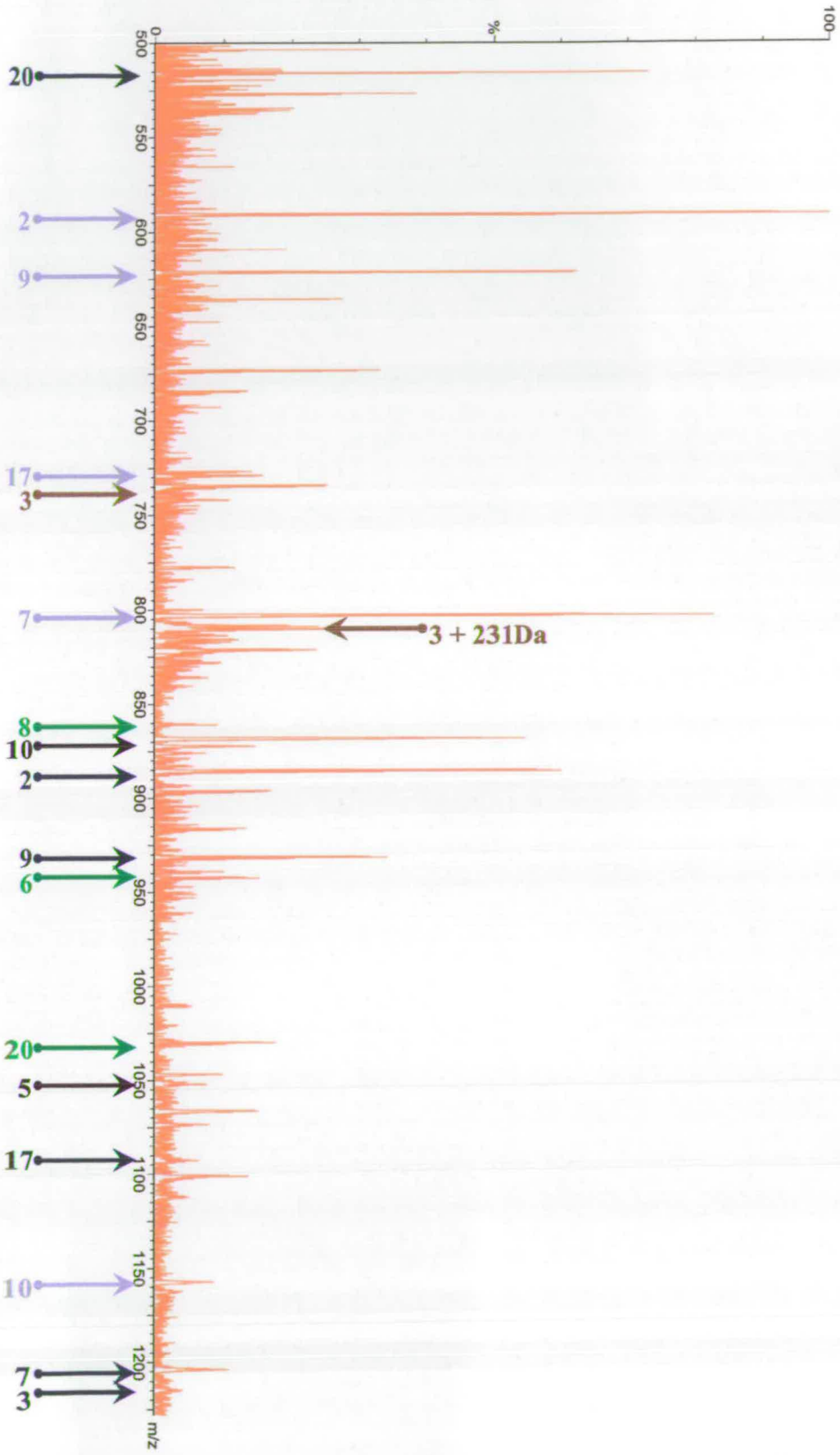


Figure 3.19b – Q-ToF MS spectra of in-gel AspN digested biotin synthase , Peptides matched with theoretical (See Figure 3.10) labelled and colour coded as to charge state as follows, M^{+1} in **green**, M^{+2} in **blue**, M^{+3} in mauve and M+4 in **black**. Peptide containing lysine 49 and corresponding +231Da species shown in **burgundy** .

The MALDI analysis displayed two large proteolytic fragments 4 and 15 of mass 6186Da and 6515Da respectively which were not observed under electrospray conditions. As in the previous analyses we clearly observed a species which would correspond to the peptide containing lysine 49, (AspN3-WT) and again a species matching the PLP-modified species (AspN3-WT+PLP). In this case both peptides were observed in their triply-charged state at 734.17. m/z and 811.18 m/z respectively as shown in Figure 3.19b.

AspN3-WT	DPRQVQVSTLLSIKTGACPE	743.17 m/z (M^{+3})
AspN3-WT+PLP	DPRQVQVSTLLSIKTGACPE-[PLP]	811.18 m/z (M^{+3})

These species were not as ideally suited for CID MS/MS sequence analysis as dissociation of a M^{+3} charge state species yields a complex dissociation spectra with M^{+1}, M^{+2} and M^{+3} fragment ions observed. Spectrum deconvolution software reduces complexity of acquired ESI continuum spectra by examining spectra, assigning charge state and finally displaying all species as singly charged. However we had no access to this software therefore assignment of a peptide with a mixture of charge states sequence manually is cumbersome. These species were of relatively high mass, 2200Da and 2431Da respectively, and were found to be difficult to fragment successfully in our instrument and therefore sequencing was found not to be possible. A degree of fragmentation was observed at high collision energies however the signal significantly reduced on the ToF detector and therefore insufficient ion statistics were obtained for the fragment species to enable interpretation.

For these reasons we concentrated on the MS/MS sequencing of the Lysine 49 containing peptide of interest, (Tryp4-WT), and the postulated PLP modified derivative (Tryp4-WT+PLP). MS/MS fragmentation was performed using the doubly charged peptides derived from biotin synthase digestion with trypsin with peaks observed at 608.73 m/z and 724.24 m/z respectively and is shown in Section 3.5.7.

3.5.6 – Q-ToF-MS Comparison of Residue 49 Containing Peptides of WT, K49Q and K49R Biotin Synthase.

Thus far we have investigated the possible PLP binding site of biotin synthase, lysine 49, by peptide mass fingerprinting and observation of a species that could correspond to a PLP modification. The assumption that this Lysine 49 residue is modified was further investigated by site-directed mutagenesis at this residue. The proposed PLP binding lysine was mutated to both arginine and glutamine to disrupt the proposed binding site and remove possible internal aldimine formation. These mutants should not be modified by PLP and enzymatically-derived peptides should provide diagnostic mass shifts for identification of the peptide containing the modified lysine residue. For this analysis mutant proteins were provided by Dr. Ollagnier-de Choudens prepared identically to the wild-type enzyme i.e. all samples had been purified and treated with PLP followed by $NaBH_4$ using the published method (92, 91).

These samples were in-gel digested and treated in the same way as the wild-type enzyme and digestion mixture was desalted for nanospray Q-ToF-MS analysis (as described in Section 3.5.3). Samples were analysed under equivalent conditions with mutant samples and wild-type analysed in parallel and wild type spectra were used as reference.

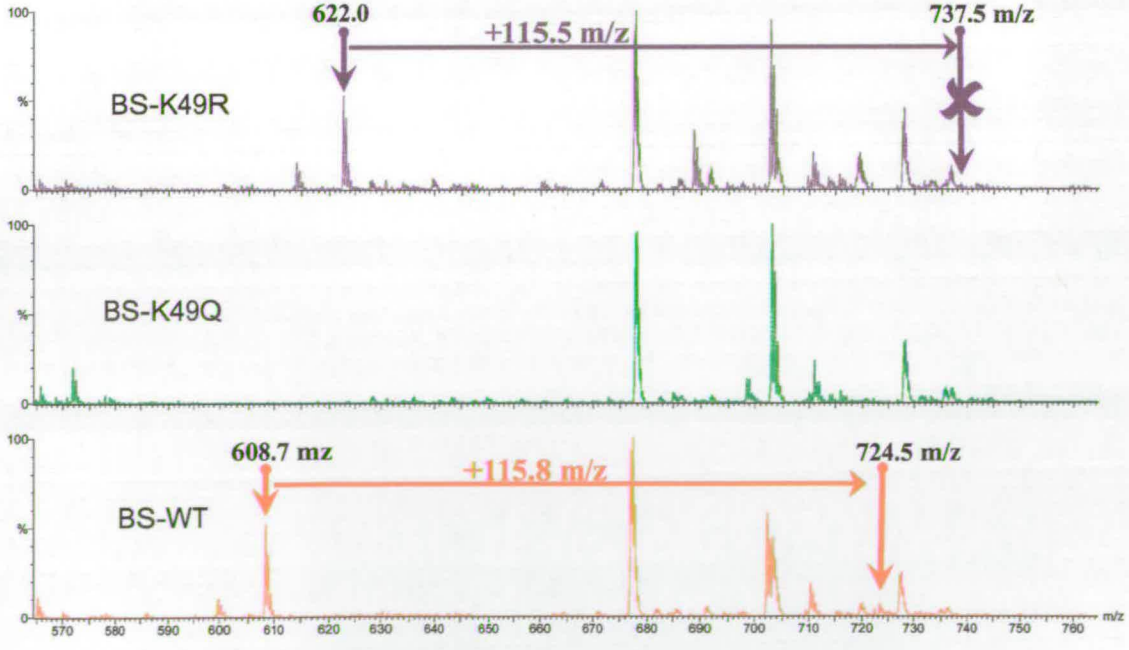


Figure 3.20 – Pile up 550 to 765 m/z region of Q-ToF spectra observed using WT (lower in red), K49Q (centre in **green**) and K49R (upper in **purple**) proteins digested with trypsin. Showing Tryp4-WT and Tryp4-WT+PLP species present in WT and absent in both mutants.

The lower trace (in Red) in Figure 3.20 shows the region 500 to 800 m/z of the Q-ToF fingerprint spectra (Figure 3.19a) of trypsin digested WT BS containing the M^{2+} (Tryp4-WT at 608.73 m/z) species of the unmodified peptide containing K49 and its corresponding modified species Tryp4-WT+PLP, seen at 724.245 m/z. The additional 231Da modification is observed as a +115.5 m/z shift due to the double charge state of the species. The centre trace is the same region of the K49Q BS fingerprint spectra (in green) which shows an absence of any species at 608 m/z , due to the lack of a trypsin cleavable residue, lysine or arginine, at position 49. Trypsin digest of mutant BS K49Q should now result in a peptide of mass 2234.5 Da (combination of trypsin peptides 4 and 5) which we have thus far been unable to observe via ESI or MALDI MS. However DNA sequencing shows the mutation has been successful. More importantly, the Tryp4-WT and Tryp4-WT+PLP peptides observed at 608.73 and 724.5 m/z in the wild-type digest are absent from the mutant spectrum. This provides greater confidence in the assignment of these peptides in the wild-type digest.

The upper trace (in purple) displays the region 500 to 800 m/z also and was acquired using the trypsin digested K49R mutant. The absence of a species at 608.73 m/z was as expected as the lysine to arginine mutation results in a 28 Da mass increase of peptide 4. This mutation is confirmed by the appearance of a new species with a m/z increase of 14 m/z at 622.m/z as indicated. To rule out the possibility that Arg could somehow be modified in a similar fashion to Lys we examined the spectra for an additional species. If Arg was modified with PLP a species would be observed at 737.5 m/z (Tryp4-K49R+PLP (622 +115.5 m/z)) and no evidence to suggest that this or another species is present in any of the acquired data.

A similar comparison using AspN digested biotin synthase analysed with Q-ToF-MS is shown in Figure 3.21. The lower trace is Asp-N digested WT biotin synthase (shown in red) and the region 700 to 850 m/z of the spectra is shown. The M^{+3} unmodified species (at 734.17 m/z) of AspN3-WT contains the proposed PLP modification site K49 (See Figure 3.10a). A species corresponding to AspN3-WT+PLP is observed in the +3 charge state at 811.18 m/z. The difference between the species is shown as +77 m/z shift consistent with the triple charged nature and equates to the +231 Da increase expected from an addition of a PLP moiety to a

peptide. The centre trace in Figure 3.21 shows the 700 to 850 m/z region of the AspN digested K49Q BS mutant spectra (in green). Since lysine and glutamine differ in mass by only 0.03638 Da the mutation of lysine 49 to a glutamine results in a mass change to the derived peptide too small to identify unambiguously. This results in the appearance of a peptide with the same nominal mass of 2200 Da that is visible in the spectra of the K49Q mutant peptide which is equivalent to the WT. This peptide is observed as a triply-charged species at 743.17 m/z and indistinguishable from the WT in this analysis.

However the AspN3-WT+PLP observed in the WT BS at 811.18 m/z is absent in K49Q digest which again gave us confidence that we had assigned the identity of this peptide species correctly. The upper trace shows the 700 to 860 m/z region of the Asp-N digested K49R BS spectra (in purple). Peptide 3 shows a species with the expected mass increase of 28 Da associated with a lysine to arginine mutation as shown and is demonstrated by a m/z shift of 9 due to the +3 charge state of the species.

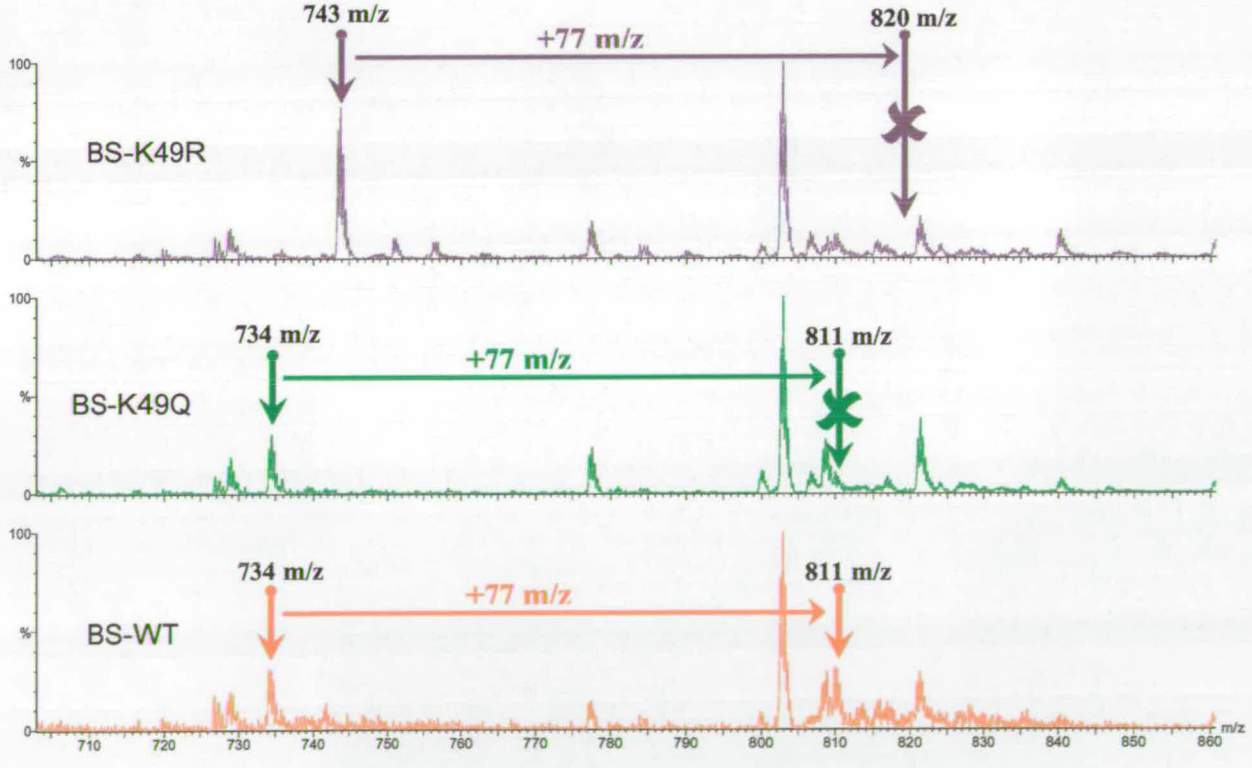


Figure 3.21 – Pile up 700 to 860 m/z region of Q-ToF spectra observed using WT (lower in red), K49Q (centre in **green**) and K49R (upper in **purple**) proteins digested with AspN. Showing Peptide3 and Peptide3+231Da species present in WT and absent in K49Q and K49R mutants.

Similarly to the trypsin-digested K49R BS mutant no species was observed with a mass shift of 231 Da relative to the unmodified peptide, indicating that the K49Q mutant did not contain a PLP-modified lysine. Combining all of the mutagenic and MS data leads us to conclude that the initially observed species that was assigned as the K49-containing fragment in the trypsin digest of WT BS is the site of PLP modification.

3.5.7 – Q-ToF based MS/MS peptide sequencing of modified and unmodified lysine49 containing peptide.

The Q-Tof MS/MS analysis was performed using identical conditions and sample preparation as for the MS fingerprint spectra. The species isolated for sequencing were the doubly charged 608.8 m/z unmodified species and the 724.2 m/z species, Tryp4-WT and Tryp4-WT+PLP respectively, of the BS wild-type tryptic digest (See figure 3.10b). These species have been subjected to CID MS/MS sequencing specifically due to the suitability of these species with regard to apparent mass, abundance and their double charge which provides a clear dissociation spectra for de-novo sequence interpretation.

The species under study was isolated in the mass spectrometer with the quadrupole of the Q-ToF set at 0.5 Da above the monoisotopic mass to enable transfer of monoisitopic 1st, 2nd and 3rd isotope peaks. Dissociation was induced using Argon as the collision gas and the voltage in the collision cell was optimised between 15V and 30V to yield maximum dissociation peak intensity and fragmentation data. When optimised for dissociation data was acquired for up to 30min or until the sample was consumed.

Figure 3.22 shows an expansion of the region containing both the unmodified species (Tryp4-WT labelled **A**) and +231 modified species (Tryp4-WT+PLP labelled **B**) of Peptide 4 and the insets are expansions of each individual species when isolated with the quadrupole before application of the Argon collision gas and collision energy (shown in Figure 3.22 - **A*/B***). When collision energy was applied to the unmodified 608.8 m/z species and data acquired the spectra shown in Figure 3.22, (**A****) was obtained. The spectra shows a single y-ion series consisting of fragments y_{10} to y_1 (indicated with - ◊) which when compared to the Biemann diagram (1)

(Figure 3.23) of the theoretical peptide sequence shows a complete complement of the dissociation fragments providing conclusive evidence for the sequence of this species being Q V Q V S T $^{I}/_{L}$ $^{I}/_{L}$ S $^{I}/_{L}$ K from N-term to C-term. It also contains a single b-ion series consisting of fragments b_2 to b_{10} (indicated with - •) and when this data is compared to the Biemann diagram (Figure 3.23) all theoretical fragments, excluding the b_1 ion at 129 m/z, were observed. In the low mass range below 130 m/z, species were observed which correspond to the immonium ions of glutamine(Q), leucine/isoleucine (L/I) and valine (V) observed at 72 m/z, 86 m/z and 101 m/z respectively. This data adds further confidence to the identification of the peptide as the theoretical Tryp4-WT derived from trypsin digestion of WT BS.

The spectra acquired when the Argon collision gas was introduced and collision energy applied to the isolated Tryp4-WT+PLP (at 724.2 m/z) species, thought to be derived from a +231 Da modification of 608.8 m/z peptide of BS, is shown in Figure 3.22 (**B****) however the acquired MS/MS spectra in this case is significantly more complex. The Spectra B** shown in Figure 3.23 includes the complimentary immonium ions also seen in the 608.8^{2+} MS/MS spectra in the low mass range and 3 distinct y series as described below. A partial b-ion series is seen (indicated with - •) consisting of the ions b_1 to b_7 but excluding the b_5 ion thus providing a sequence tag of Q-V-Q running N to C terminus from ions b_1 to b_4. Ions b_6 and b_7 provide evidence of a single L or I residue. Also observed is the partial y-ion series consisting of y_2 to y_6 and provides an N to C terminus sequence tag of T-$^{L}/_{I}$-$^{L}/_{I}$-S which also corresponds to the theoretical amino acid sequence composition. A related y-ion series is observed with the addition of a y_8 ion (therefore $y_2 - y_6$ plus y_8) which is consistent with the unmodified species already sequenced from the 608.8 m/z species. This data suggests that the +231Da moiety is cleaved from the lysine during the collision induced dissociation thus yielding an unmodified species which is subsequently further dissociated to provide these fragment ions. Despite numerous attempts to affect the cleavage of PLP we were unable to cause the loss of this modification without significant backbone fragmentation.

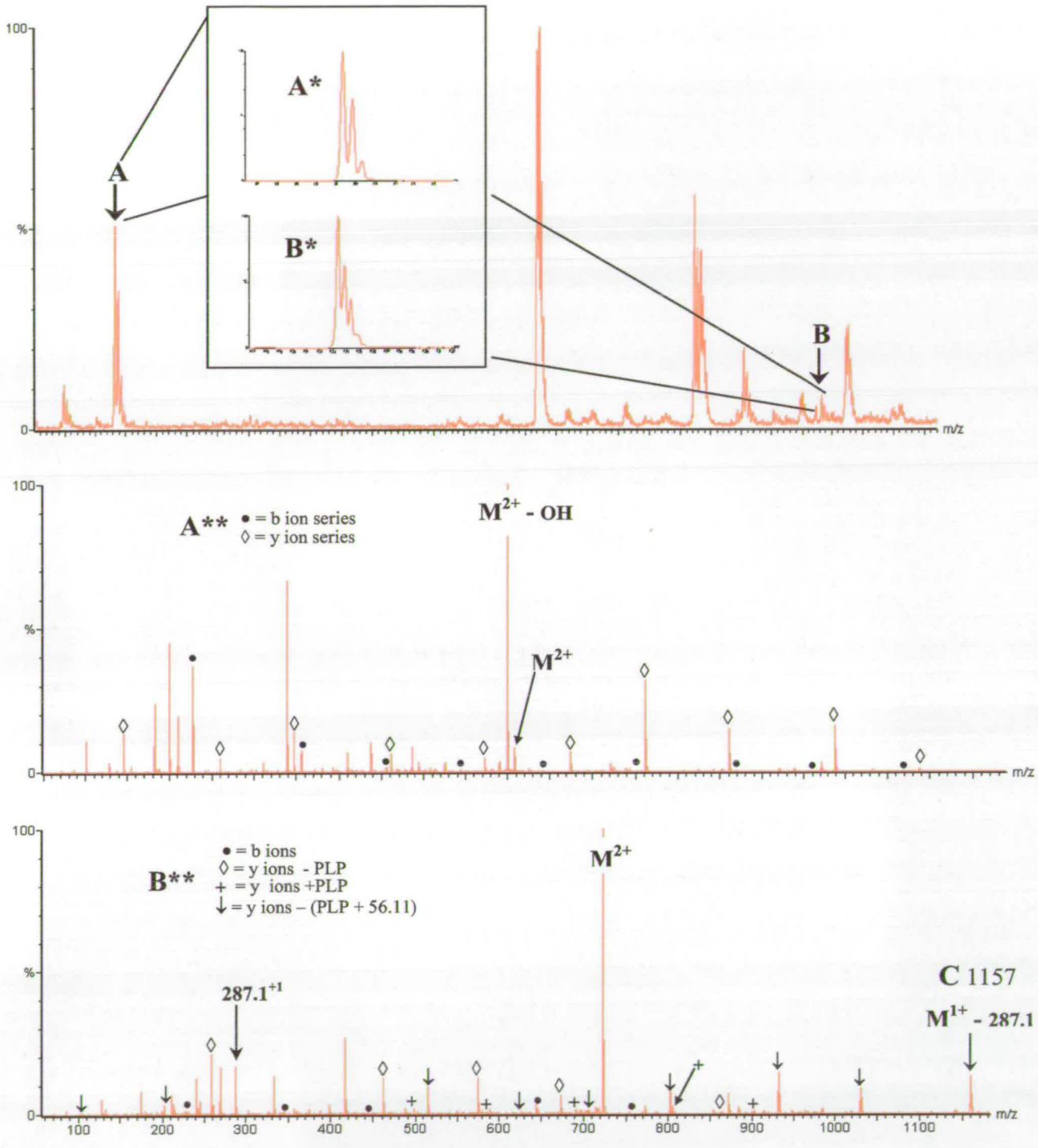


Figure 3.22 – Q-ToF MS^1 spectra showing unmodified (Tryp4-WT = **A**) and modified (Tryp4-WT+PLP = **B**) peptide containing lysine49. The inset shows the isolated peaks before introduction of collision gas (**A**/**B***). Q-ToF-MS^2 sequencing of the unmodified species (**A****) showing both y-ion and b-ion series and Q-ToF-MS^2 sequencing of +231 modified species (**B****)showing partial b-ion and y-ion series, partial y-ion series with loss of 231 Da, partial y-ion series with loss of 287 Da.

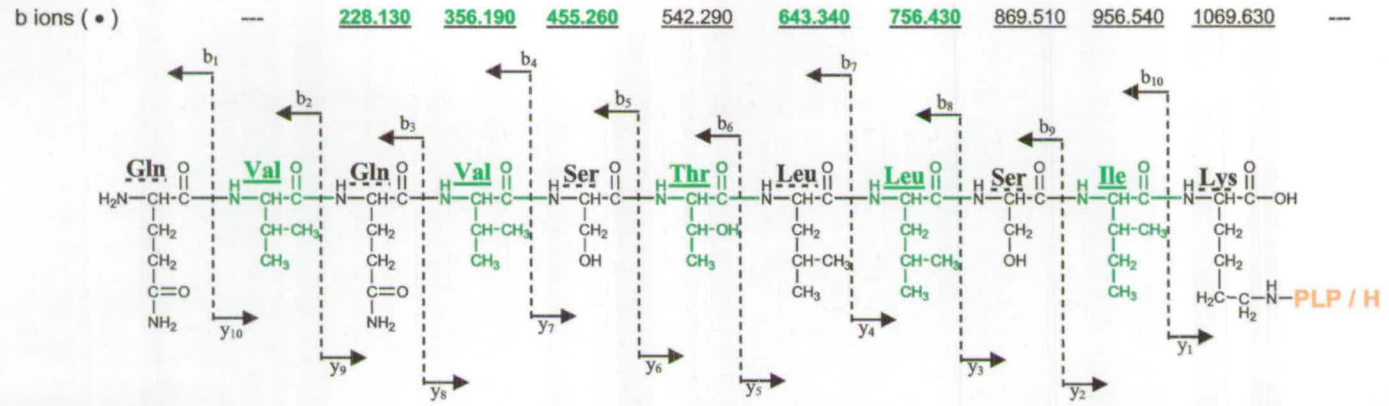
The final series we observed displayed the highest relative intensity and yielded the most complete set of y-ions as y_{10} to y_5 and y_2 and y_1 were observed. Initial examination of the data very quickly provided evidence for the N to C terminal sequence consisting of Q-V-Q-V-S-T, which perfectly matches the sequence of Tryp4-WT+PLP. However under further investigation all the species from this series exhibited a delta mass of + 287 Da from the unmodified Tryp4-WT sequence.

The appearance of a single charged species **C**, an apparent molecular ion at 1157, is consistent with the parent ion minus 287Da. From this species **C,** a fragment series was observed from which the sequence Q-V-Q-V-S-T could be read. Importantly, we also observed a single charged species at 287 m/z which is not assigned as a sequence fragment ion and originates from a single cleavage of Tryp4-WT+PLP. We wanted to determine the structure of this 287 Da ion and explain a mechanism for its formation during collision induced dissociation but did not have access to MS^3 instrumentation which may have enabled this. However, the appearance of the 287^{1+} m/z peak shows that this species must originate from a single cleavage and not as a multiple cleavage event, the cleavage of PLP and a valine sidechain (C_4H_8 = 56) for example could provide this apparent molecular ion but as we observe the 287^{1+} m/z species this must be a single fragment. One plausable explanation is that the cleavage of the PLP moiety is occurring along the site of attachment, namely the lysine sidechain and corresponds to cleavage of the beta-gamma C–C bond of the lysine side chain as shown in Figure 3.24. This cleavage, although energetically unfavourable, may be promoted due to the nature of the gas phase structure of the peptide with the PLP bound at the C-terminus. Cleavage may be aided by interactions with the backbone and C-terminus of the peptide in an electrostatic cyclisation, so stabilising the peptide bond. Another effect could be due to polarisation of the side chain and a subsequent decrease in the bond energy favouring the cleavage of the Cβ-Cγ bond over N-N linkage between K and PLP. After the cleavage of the Cβ-Cγ bond the resulting dissociation species would correspond to a species with mass 288 Da and the apparent molecular ion with mass which we do in fact observe in the MS/MS spectra. This unusual cleavage provides little or no opposition to the identification of the observed 724.2 m/z species indeed being the

enzymatically derived Tryp4-WT+PLP peptide of BS but is an intriguing observation as to the nature of this species in the gas phase.

3.5.8 – Conclusions

Through a combination of site-directed mutagenesis and mass spectrometric analysis we have been able to further investigate the proposed presence of a protein bound PLP moiety. The observation that biotin synthase had pyridoxal phosphate-dependent cysteine desulfurase activity by Ollagnier-de-Choudens *et al.*, implicates a highly conserved lysine 49 residue in *E. coli* biotin synthase. We have characterised enzymatically derived peptides with masses corresponding to the addition of PLP to this conserved residue. No PLP-derived peptides were observed when we mutated the lysine side-chain. Moreover, we successfully isolated and sequenced both the unmodified and PLP-modified peptides by gas-phase collision-induced methods. Final conclusions from this study strongly confirms the proposal that biotin synthase can be covalently modified by PLP and that this occurs specifically at the conserved lysine residue found on a loop only 7.5Å from the [4Fe-4S] cluster of the holo-protein as shown by the recently published crystal structure. Although this residue is not within a known PLP binding site observed in other PLP-dependent enzymes, the solution phase dynamic structure of biotin synthase is significantly more complex than the single 'snap shot' crystal structure elucidated by Drennan and co workers (16) as discussed in Section 1.3.5.



y ions (◊)	---	1087.670	988.600	860.550	761.480	674.450	573.400	460.310	347.230	260.200	147.110
y ions + PLP (+)	---	1318.670	1219.600	1091.550	992.480	905.450	804.400	691.310	578.230	491.200	378.110
y ions + PLP – 287 (↓)	---	1031.570	932.500	804.450	705.380	618.350	517.300	404.210	291.130	204.100	91.010

Figure 3.23 – Biemann notation (1) of theoretical CID species of Peptide4 (residues 39-49) derived form trypsin digestion of Biotin synthase with (+) and without (◊) PLP modification at lysine49. Also shown is species corresponding to a loss of 287Da from the M^{+2} molecular ion (↓) resulting in the 1157^{+1} m/z apparent molecular ion species. Species observed in unmodified peptide shown underlined and species observed in modified peptide shown in **green**.

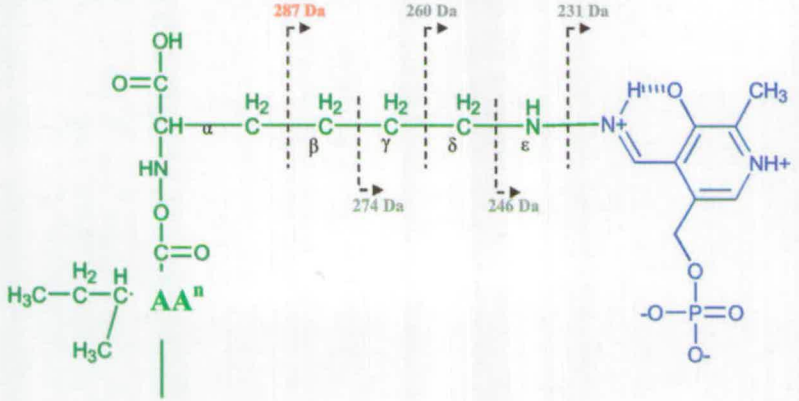


Figure 3.24 – Theoretical fragmentation of PLP modified lysine side chain and the resulting mass losses for each site. Observed loss is shown in red.

3.6 – *E. coli* Biotin Synthase as a Pyridoxal-Phosphate Dependant Cysteine Desulfurase

3.6.1 – Preparation of Pyridoxal-Phosphate loaded Biotin Synthase

The 6HisBS protein and mutants were prepared in a 100 μM PLP cell lysis solution for subsequent analysis fresh PLP was added to protein solution by dialysis of protein solution against 100 μM. Protein solution was monitored during the process, (before dialysis, after dialysis and following gel filtration cleanup) by UV-Vis scan from 250nm to 600nm. The protein aliquot was defrosted on ice and dialysed against Buffer C with addition of 100 μM PLP for 3 hours and unbound PLP removed by Gel Filtration using a PD10 disposable column from Pierce. The UV-Vis scans are shown in Figure 3.25 and maxima ratios shown in Figure 3.26. UV-Vis analysis performed prior to fresh PLP addition displays characteristics as described previously (Section 3.3) and scan performed directly after PLP dialysis clearly shows an increase in the observed maxima at 325nm and 400nm consistent with free PLP present at high concentration in solution. The most striking difference is observed after PD10 column clean up, this clearly shows an increase in the 425nm region with respect to the 280nm region. In contrast little difference is observed in the 330nm region compared to 280nm region. These increases are highlighted in Figure 3.26 which shows the calculated ratios for 330nm/280nm, 425nm/280nm and 425nm/330nm.

These simple measurements suggest that the addition of fresh PLP to the protein solution dramatically shifts the absorbance characteristics of the solution which would lead to the conclusion that biotin synthase is effected by the presence of PLP and that this may have a structural or chemical role.

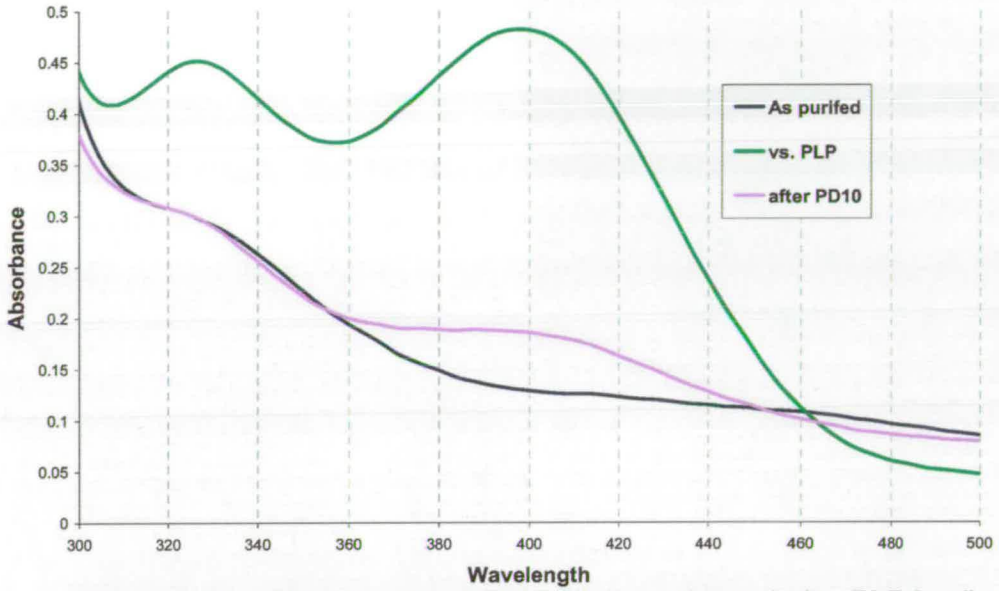


Figure 3.25 – UV-Vis scan of *E.coli* biotin synthase during PLP loading of protein. Before dialysis (blue), after dialysis against a PLP-containing buffer (100 μM PLP) (green) and after re-purification using PD10 GF column (pink).

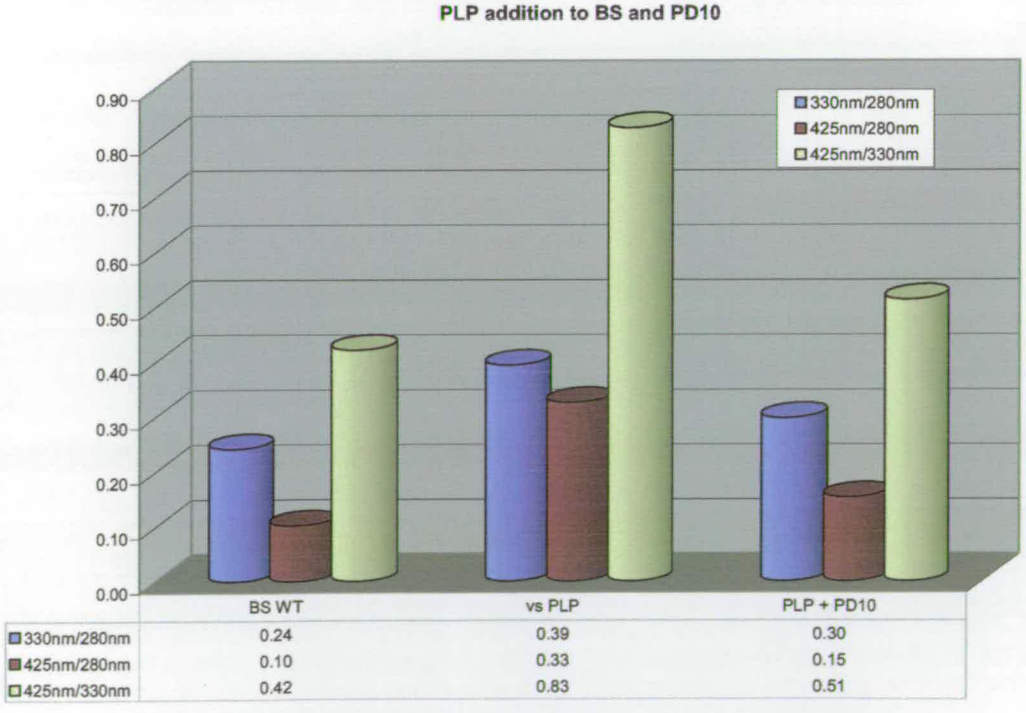


Figure 3.26 – Bar chart of UV-Vis maxima ratios of 280, 330 and 425 nm before and during PLP loading.

3.6.2 – Incubation of PLP loaded BS with L-Cysteine Monitored by UV-Vis Spectroscopy

The UV-Vis analysis of 6His-BS incubation with cysteine was optimised using protein at approximately 1 mg/ml and incubation at 37 °C. A maximum response was observed when cysteine was added to the protein sample at 3 mM final concentration. Incubation of the protein sample final concentration of cysteine greater than 3 mM resulted in the mixture immediately turning turbid and thus making UV-Vis analysis impossible.

The spectrophotometer was blanked using buffer (50 mM Tris-HCl, pH7.5) and 3 mM cysteine final concentration. Identical conditions were used to assay the BS-containing sample. Addition of cysteine immediately caused absorbance of the sample to increase across the region between 250 nm and 600 nm. This initial absorbance increase could be caused by the initial mixing. After 30 sec the absorbance fell to approximately the initial values. As shown in Figure 3.27 after 10min incubation at 37 °C the absorbance at 330 nm had increased (A) and absorbance at 425 nm had decreased (B). This trend showing an increase at 330 nm and a decrease at 425 nm is continued throughout the experiment until time 40 min after which no measurable change was observed. The increase at 330 nm and decrease at 425 nm converge on an isosbestic point at approximately 360 nm, as shown in Figure 3.27, labelled (C). The appearance of an isosbestic point in a UV-Vis spectrally monitored time course is indicative of the chemical conversion of two species. A well characterised cysteine desulfurase, NifS, has been shown to respond to addition of L-cysteine with similar spectral characteristics and these observations were suggested to be indication of conversion of the NifS-PLP internal aldimine to the NifS PLP-L-cys external aldimine (145,172).

When K49Q and K49R biotin synthase mutants were analysed in an identical experiment, the addition of L-cysteine affected the same initial increase from 250 nm to 600nm (with return to initial values within 30sec). However after a 40 min incubation period at 37 °C no distinct absorbance changes were observed in the spectra. This result tends to suggest the involvement of the lysine residue 49 in our observed cysteine activated absorbance changes. As suggested a mutation at this site

would hinder incorporation of PLP into the protein structure and thus affect the proposed cysteine desulfurase activity of the protein.

Incubation of the BS C53/57S and BS C53/60S mutants with 3 mM cysteine at 37°C caused the comparatively unstable mutants to precipitate thus making the analysis impossible. When these double mutants were incubated for the 40 min time course with L-cysteine, final concentration of 2 mM, they showed an absorbance profile much the same as for the wild type enzyme. The increase at 330 nm and decrease at 425 nm were observed albeit much less pronounced. The reduced observed change caused by L-cysteine addition may be a direct result of the reduced cysteine concentration. However, as protein concentration is in the micro molar range compared to the milli molar range for PLP this seems unlikely. A more likely explanation is that the reduction is indicative of more fundamental changes to the protein structure caused by the mutation of the cysteine residues.

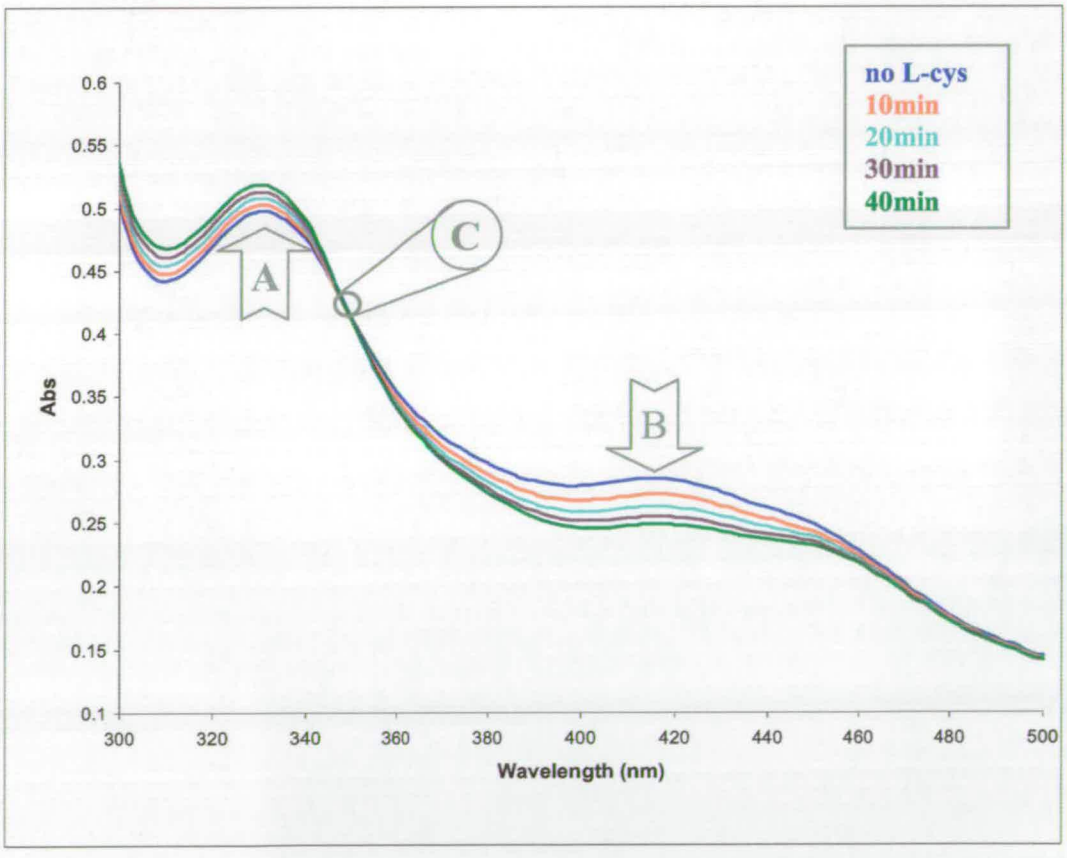


Figure 3.27 – Incubation of wild-type BS (26μM) re-loaded with PLP (excess PLP removed by gel filtration) with 3mM L-cysteine. A and B indicate the increase and decrease in absorption at 330nm and 425nm respectively during the 40min incubation period. The isobestic point produced by these spectral changes is seen at approximately 360nm and is labelled C.

3.6.3 – Conclusions

We have shown that the addition of PLP has a marked effect on the UV-Vis characteristics of the protein which suggests the protein has the ability to bind PLP. This PLP-binding is entirely dependant on the presence of lysine49 in the protein. As-purified BS has a characteristic absorbance spectrum due to the presence of Fe-S clusters and incubation of PLP causes small changes in the small spectrum. Incubation of this PLP-reloaded BS with L-cysteine caused absorbance changes such that after prolonged incubation the BS resembles that of as-purified biotin synthase. The changes caused by addition of L-cysteine closely resemble those previously observed for cysteine desulfurase enzymes (e.g. NifS) which have been attributed to the conversion of an enzyme-PLP internal aldimine to the corresponding PLP-cysteine external aldimine (145,172). These observations lead us to conclude that our preparation of BS contains a species that is capable of binding PLP. In addition, this PLP binding species can subsequently be converted to a PLP-cysteine aldimine. The amount of the PLP-binding form of BS in the total BS sample is unknown and this would be difficult to determine. It is now accepted that turnover of BS destroys at least one of the Fe-S clusters, which then has to be regenerated. The role that the PLP-bound form of BS plays is the subject of controversy but we speculate that it could be involved in the delivery of sulfur to the partial cluster allowing regeneration of this to the active form.

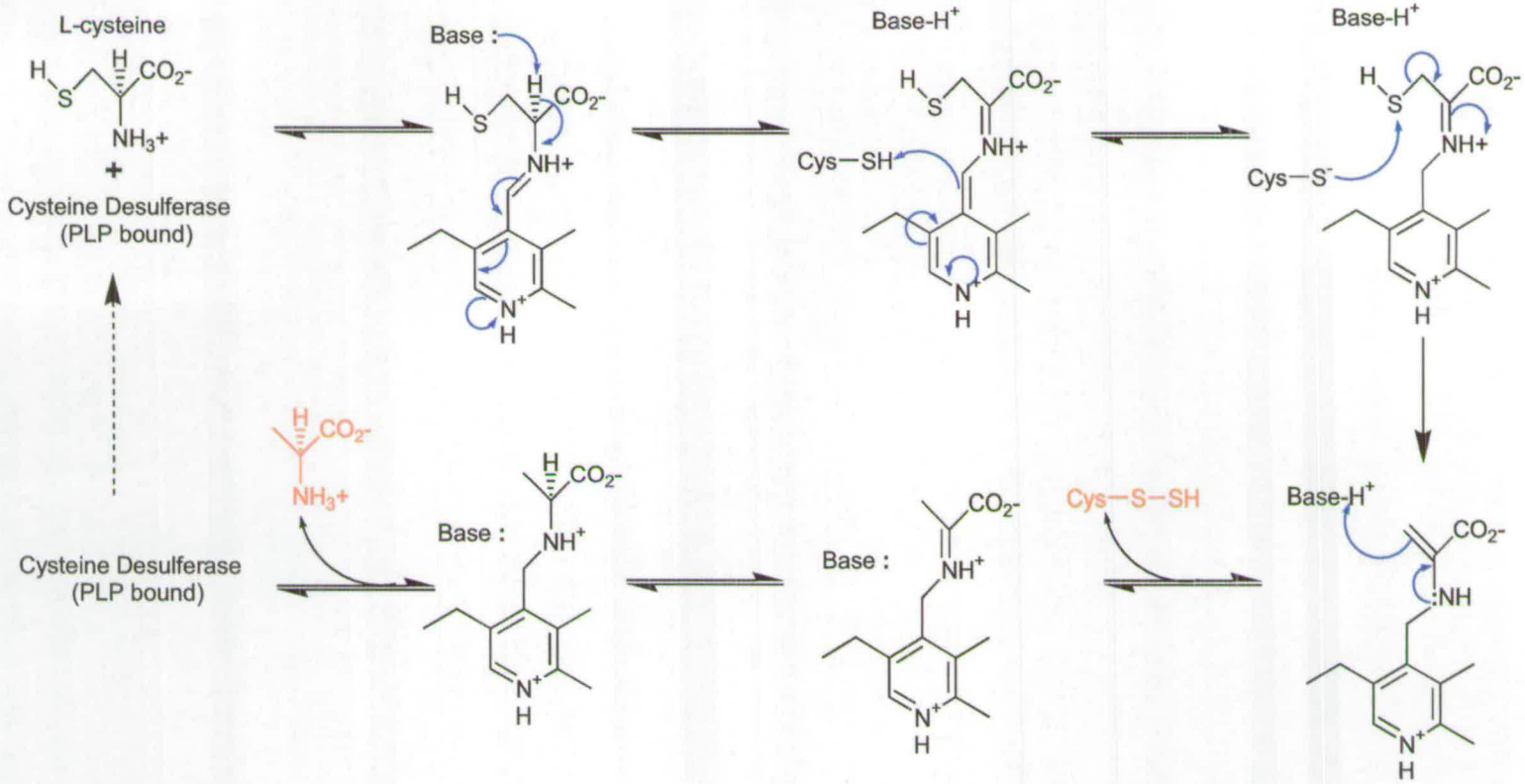


Figure 3.28 – Hypothetical mechanism of activity of a PLP dependant cysteine desulfurase enzyme (NifS, biotin synthase?). The mechanism is dependant on availability of PLP, L-cysteine, active site base (lysine) and a S acceptor. Modified from Zheng *et.al.*(6)

3.7 – Differential Expression Analysis of Biotin Synthase by Two Dimensional Gel Electrophoresis

3.7.1 – Introduction

As discussed in Section 1.2.7 the *in vitro* biotin production from dethiobiotin (DTB) using the standard conditions results in only approximately one conversion of DTB to biotin. It has been postulated that an explanation for this could be that a important component or components (protein, small molecule, peptide etc) are missing from the *in vitro* assay, the addition of which may result in enzymatic turnover. The identification of these missing components is of great interest with the goal of producing a fully catalytic enzyme *in vitro*. This would help in our further understanding of the complex bio-reaction and possibly solve several of the mechanistic questions, which as yet are unanswered. Although several protein partners, such as flavadoxin (FLD), flavadoxin reductase (FLDR) and small molecules, such as SAM and NAPDH, have been shown to enable the single turnover of biotin synthase to convert DTB to biotin, it is still a contentious issue as to their *in vivo* role. The combination of these species certainly seem to provide a resource for activity *in vitro,* however thus far no hard evidence has been amassed to unequivocally prove they are the true partners in the cellular environment. Indeed, the lack of catalytic turnover points to two conclusions, they are true catalytic partners and we are missing an important component of the system or, the currently observed single turnover is an artefact of the assembled multi component system.

The production of the 5'-deoxyadenosine radical ($DOA^{\bullet}$) is vital to the activation of the carbon 6/9 of DTB. This radical is thought to derive from reductive cleavage of SAM by a reduced [Fe-S] cluster formed via a cascade reduction involving NADPH, FLDR and FLD, in this way the activity is dependant on the existence of the iron-sulfur cluster. We postulated that among other possible protein partners the cluster will have to be formed sequentially for each turnover by some means.

As the ISC operon and associated proteins have been shown to be vital for [Fe-S] cluster formations and purified biotin synthase contains a preformed [Fe-S] cluster and as over expression of biotin synthase has been shown to increase biotin production we hoped to observe that a forced over-expression of biotin synthase could up regulate cellular production of the *in vivo* protein partners involved in its cluster formation as well as those involved in DTB to biotin conversion. We attempted to analyse these up-regulations by differential expression monitored using 2Dimensional Electrophoresis as described below.

3.7.2 – Preparation of Extracts for 2Dimensional Electrophoresis

The differential analysis cell growths were produced by small scale (10ml) expression (Sections 2.16) using pET16b/BioB and pET16b as shown in Figure 3.29 and using BL21(DE3) competent cells. Cell extracts prepared by incubation of cell pellet in 8M urea and CHAPS detergent to facilitate the solubilisation of the majority of the cellular proteins, full method is described in section 2.23.1.

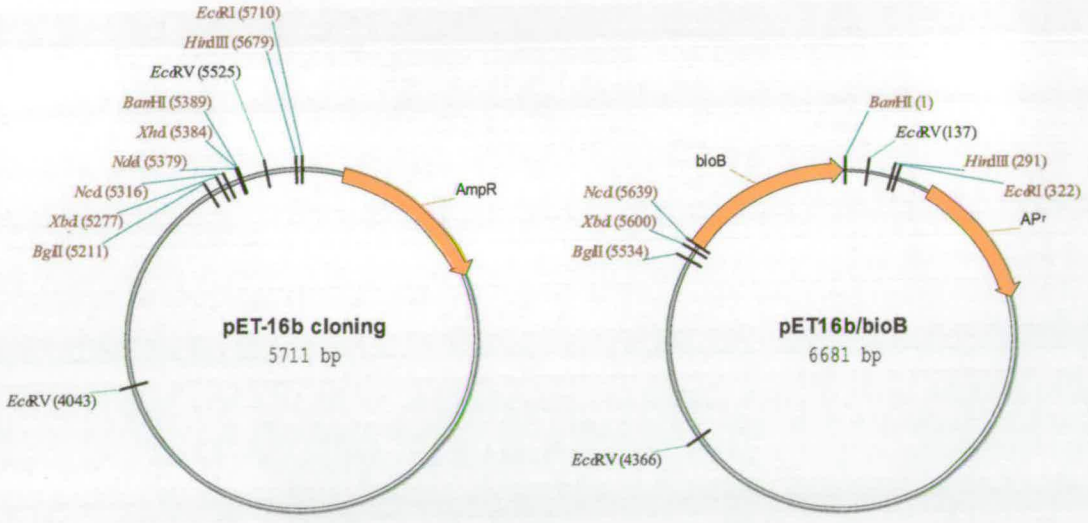


Figure 3.29 – Plasmids maps of pET16b and pET16b/bioB used for differential expression and extracts used for subsequent 2Dimensional Electrophoresis.

The cell extracts were first analysed by 1D SDS-PAGE to gauge the extent of expression and concentration of total protein for subsequent 2D analysis as shown in Figure 3.30. The biotin synthase band is clearly observed in lane 2 at approximately 40kDa as judged by the low molecular weight marker (Amersham Pharmcia) shown in lane M.

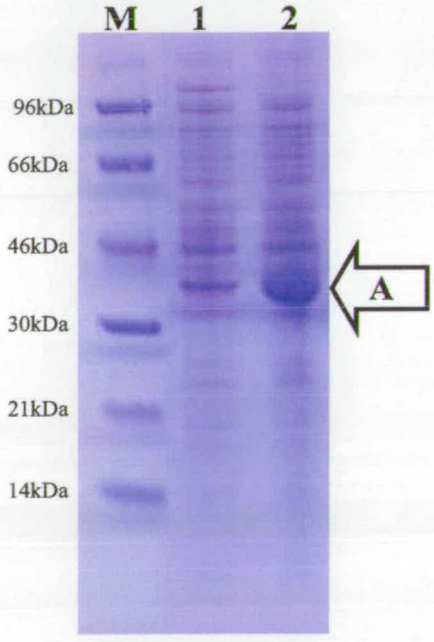


Figure 3.30 – Nu-PAGE gel showing urea solubilised cell extract for pET16b (lane1) and pET16b/bioB (lane2) expressing BL21(DE3) cells. **A** – Biotin synthase at approx. 40kDa.
M – molecular weight marker (LMW, Amersham Pharmacia) [96kDa = Phosphorylase b, 66kDa = BSA, 46kDa = ?????????, 30kDa = Carbonic anhydrase, 21kDa = Trypsin inhibitor, 14kDa = Lysosyme].

3.7.3 – Analytical 2Dimensional Electrophoresis

In the first instance to maximise coverage of the analysis we decided to perform a selection of analytical two dimensional gels which involved a wide pH range for the first dimensional separation from pH3 to pH10. It is vitally important to optimise load amounts by using dilutions of the prepared cell extract to enable visualisation of the greatest dynamic range but also to minimise streaking caused by overloading the gel. The experiment was carried out in duplicate for cells expressing biotin synthase and cells transformed with the pET16b plasmid with no insert as described above. Visualisation of protein spots was done using the PlusOne Silver Stain Kit (Amersham Pharmacia) which is not compatible with enzymatic digestion and extraction, due to the inclusion of glutaraledhyde which crosslinks the embedded protein to the gel matrix. The inclusion of glutaraldehyde maximises the sensitivity of the stain to enabling visulisation of 0.2 ng of protein per spot. This high stain sensitivity coupled with the wide pH range allowed the greatest coverage of the protein mixture for this analytical stage. Figure 3.31 shows the acquired 2-Dimensional gel for the control cell extract of pET16b transformed cells only and Figure 3.32 shows the 2Dimensional gel for the biotin synthase expressing cell extract of pET16b/bioB transformed cells.

The expressed biotin synthase can clearly be observed at approximately 40kDa but shows significant streaking across the pH range 3 to 5. A significant proportion of this streaking is certainly due to the quantity of biotin synthase loaded but even with this relatively low resolving PI strip several clearly defined biotin synthase spots are observed.

Initial visual comparison can easily identify patterns of similarity between the 2 gels and so it was reasonable to believe that this data would be suitable of the more rigorous computer based differential mapping analysis as described below.

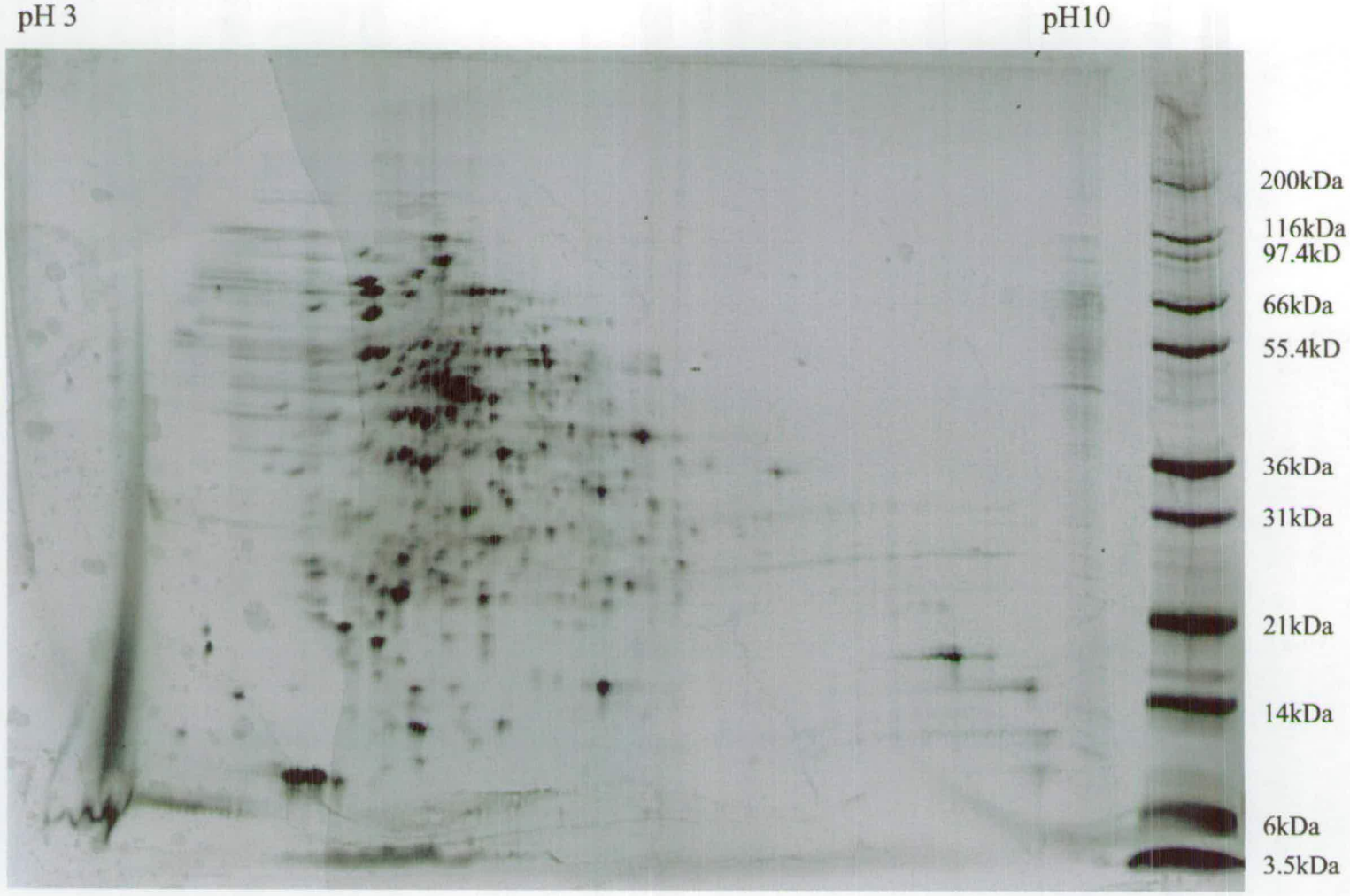


Figure 3.31 – Analytical 2 dimensional electrophoresis Gel of Urea extract from BL21(DE3) cells transformed with pET16b plasmid with no insert. 12cm pH3-10 IPG strips used as first dimension and 0.5mm 4-12% Excel Gel used for SDS-PAGE second dimension. Non Mass Spectrometric compatible silver stain (PlusOne, Amersham Pharmacia) used to visualize protein bands. Image captured and manipulated with HT-Analyser (Genomics Solutions).

Figure 3.32 – Analytical 2 dimensional electrophoresis Gel of Urea extract from BL21(DE3) cells transformed with the pET16b/bioB plasmid and expressing biotin synthase. 12cm pH3-10 IPG strips used as first dimension and 0.5mm 4-12% Excel Gel used for SDS-PAGE second dimension. Non Mass Spectrometric compatible silver stain (PlusOne, Amersham Pharmacia) used to visualize protein bands. Image captured and manipulated with HT-Analyser (Genomics Solutions).

3.7.4 – Differential analysis of 2D Gels using HT-Analyser.

The software used to analyse all 2D gels was the Investigator HT Analyser V.2.1 which involves a combination of automatic and manual matching of spots from both the control and the experimental gel. Although all precautions are taken to ensure the gels are run under the same conditions minor dimensional differences inevitability occur. Because of this to facilitate matching gel warping is used. Similar spots from the 2 gels are matched manually which allows software to warp the gel dimensions and as such identify other species with significant overlap between the 2 conditions. Although the software allows fully automated spot identification and matching, we found that a large amount of manual correction was required to eliminate false matches. For this analysis the automated spot matching functions were disabled and spot identification and matching performed manually using 'gel warping'. Shown in Figure 3.33 and Figure 3.34 is the biotin synthase expressing gel highlighting matched and unmatched spots respectively.

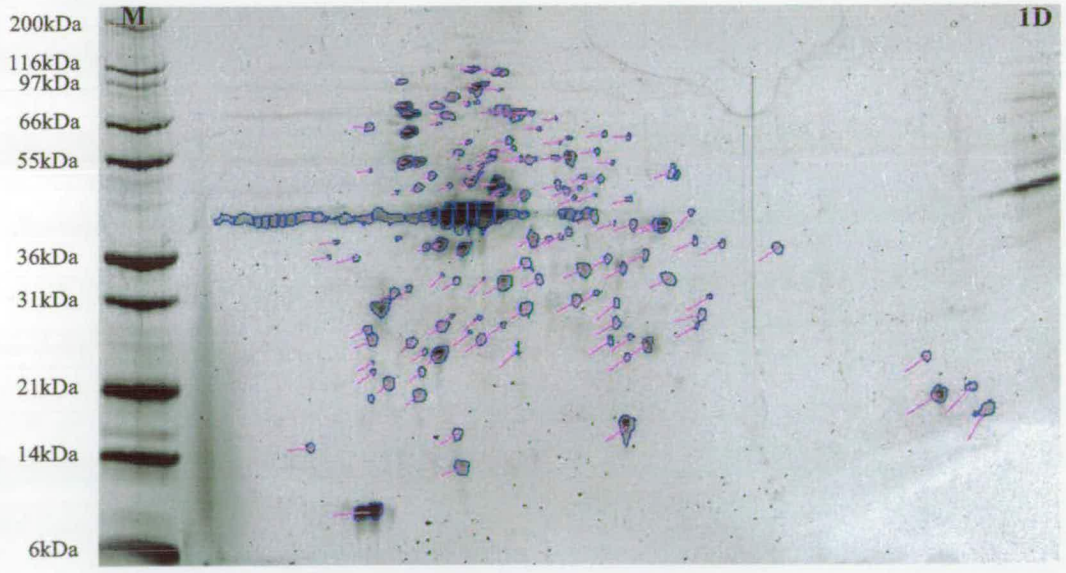


Figure 3.33 – Spots matched (shown **blue** outline with pink vector) to control gel generated from comparison of Figure 3.31 and 3.32. Vector (pink) shows relative position shift of matched control spot. Image captured and manipulated with Investigator HT Analyser (Genomics Solutions). **1D** – sample of cell extract run 2nd dimension only. Labelled regions referred to in discussion. **M** – molecular weight marker (Mark12, Invitrogen) [200kDa = Myosin, 116kDa = α - galactoside, 97kDa = Phosphorylase b, 66kDa = BSA, 55kDa = Glutamic dehydrogenase, 36kDa = Lactate dehydrogenase, 31kDa = Carbonic anhydrase, 21kDa = Trypsin inhibitor, 14kDa = Lysosyme, 6kDa = Aprotinin, 3,5kDa = unresolved Insulin]

From this analysis we were able to identify an area of the gel which shows several species which are completely absent in the control gel. Our assumption is that these are directly associated with the over expression of biotin synthase within the cell.

The majority of spots observed on the gel where the cell line is expressing biotin synthase were able to be matched directly to equivalent spots on the control gel as shown in Figure 3.33, however many of the matches showed significant intensity decreases which may have arisen due to the expression of biotin synthase but this proved difficult to ascertain due to general protein concentration reduction seen as a result of protein over expression. In a similar experiment this phenomenon was observed where the *bio*H gene product was over expressed and cell extract compared to control sample, both were analysed by 2D Electrophoresis.

Figure 3.34 highlights spots and regions (A to E) which could not matched to any observed species on the control gel. Spot A is observed at approximately 95-90 kDa and although not matched is of very low intensity. This particular species is not distinguishable in the control gel and this is most likely due to the streaking that occurred in this region resulting in the combination of two species. The large region B clearly shows the biotin synthase protein species at approximately 40kDa as expected however this species shows a huge horizontal streak from approximately pH3 to pH6. Whereas a large majority of this streak is due to the quantity of expressed biotin synthase resulting in overloading of the strip in this region, several significant well defined species lie within the region of pH 5 to 6.

This suggests that the *in vivo* expression of biotin synthase results in a number of different species with various defined pI's. The theoretical pI of biotin synthase as estimated examination of the primary sequence is 5.8. This gel suggests that this is a reasonably accurate it is clear that several distinct species are present in this preparation. Spot C is of apparent molecular weight of approximately 50kDa and is an irregular shape and is of very low intensity, it is likely that the identification of this area as a protein spot is incorrect and colour intensity in this region is due to irregularity of image while captured. Spots D and E seem to be have high intensity and regular shape consistent with regular protein species. Spot E seems to be a single species whereas the

area highlighted in D appears as if it consists of at least 3 species which have not been fully resolved in the gel.

From this initial analysis it seems reasonable to focus on the identification of the species found in areas D and E as well as confirming the identity of the species in area B assumed to be biotin synthase.

With our targets clear we repeated the analysis to include the following variations, firstly we narrowed the pH range of the first dimension and the gels were silver stained using a modified protocol compatible with Mass Spectrometry. The Immobilized pH gradient strips were changed from the pH 3-10 range to pH 4-7 in an attempt to resolve the mixed species shown in area D of Figure 3.34. The modified Silver Stain protocol excludes Glutaraldehyde and Formaledhyde which are used to fix protein to the gel matrix thus preventing extraction of peptide species from the gel slices.

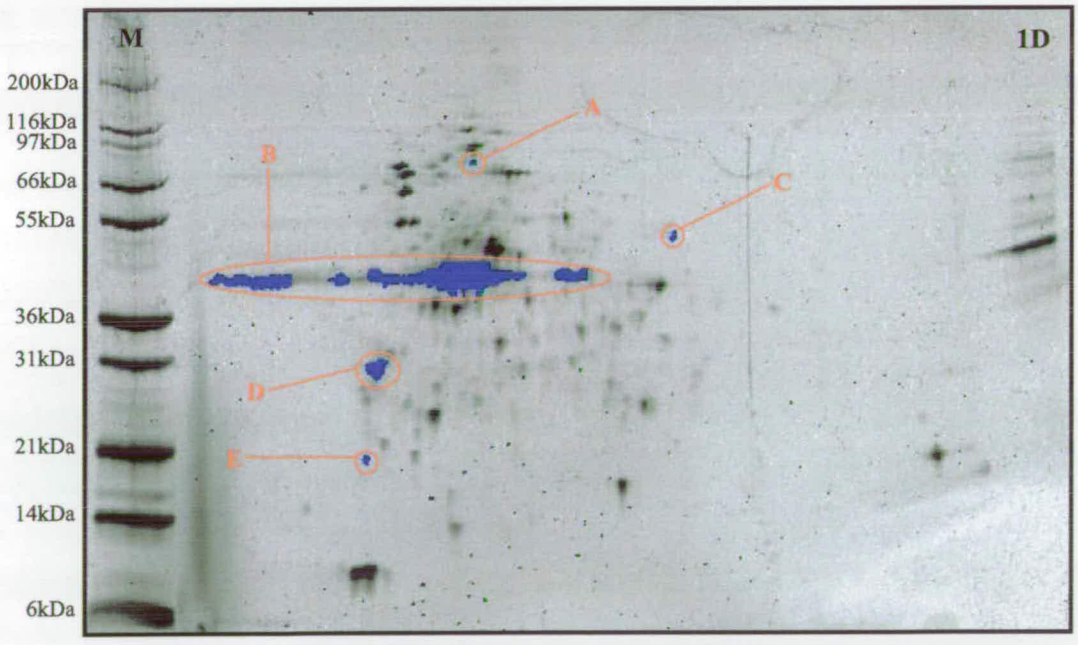


Figure 3.34 – Spots unmatched (shown **blue**) to control gel generated from comparison of Figure 3.31 and 3.32. Image captured and manipulated with Investigator HT Analyser (Genomics Solutions). **1D** – sample of cell extract run 2nd dimension only. Labelled regions referred to in discussion. **M** – molecular weight marker (Mark12, Invitrogen) [200kDa = Myosin, 116kDa = α - galactoside, 97kDa = Phosphorylase b, 66kDa = BSA, 55kDa = Glutamic dehydrogenase, 36kDa = Lactate dehydrogenase, 31kDa = Carbonic anhydrase, 21kDa = Trypsin inhibitor, 14kDa = Lysosyme, 6kDa = Aprotinin, 3,5kDa = unresolved Insulin]

3.7.5 – Spot Picking and Protein Identification by MALDI-ToF-MS and Database Querying

As discussed the species outlined above were our initial targets for identification as our analysis suggests that these species are directly associated with the *in vivo* production of biotin synthase, in particular the region D from Figure 3.34 which corresponds to species 5 to 13 in Figure 3.35. As well as these species we also attempted to identify some of the higher abundance proteins present in both control and biotin synthase expressing gels to act as internal corroboration of our experimental procedure. Shown in Figure 3.35 is the gel resulting from a biotin synthase expressing cells using the truncated pH gradient of pH 4-7, spots highlighted and labelled are gel regions excised for in-gel digestion with trypsin and subsequent MALDI-ToF-MS analysis and database searching.

All excised gel portions were treated and digested with trypsin as described in Section 2.24.1 samples were desalted using C18 ZipTips (Millipore) as manufacturers instructions and analysed by MALDI-ToF-MS as described in Section 2.29 using CHCA matrix (Section 2.2). Spectra were analysed by DataExplorer software (Applied Biosystems) resulting in a monoisotopic mass list internally calibrated with observed trypsin autolysis peaks which was used to query the SwissProt and OWL non-redundant databases (all species). Databases were queried using ProteinProspector (http://prospector.ucsf.edu/) and Mascot (http://www.matrixscience.com/) search engines and results compared for consistency and to increase validity of identification.

Excised areas of gel that produced valid database hits are highlighted in **light green** and non-valid hits are highlighted in **dark green** in Figure 3.35 and table in Figure 3.36 below shows all identified species and their corresponding database hits found from both ProteinProspector and Mascot search engines.

Highlighted sections of the gel 5, 6 and 11 to 14 when analysed produced no significant hits corresponding to *E. coli* proteins with molecular weight within the region of the gel that the protein spot was visualized. These regions varied in spot intensity but the species of greatest intensity with which we failed to identify a protein was spot 14. MS fingerprint data from species 1 and 2 both showed extremely high statistical scores

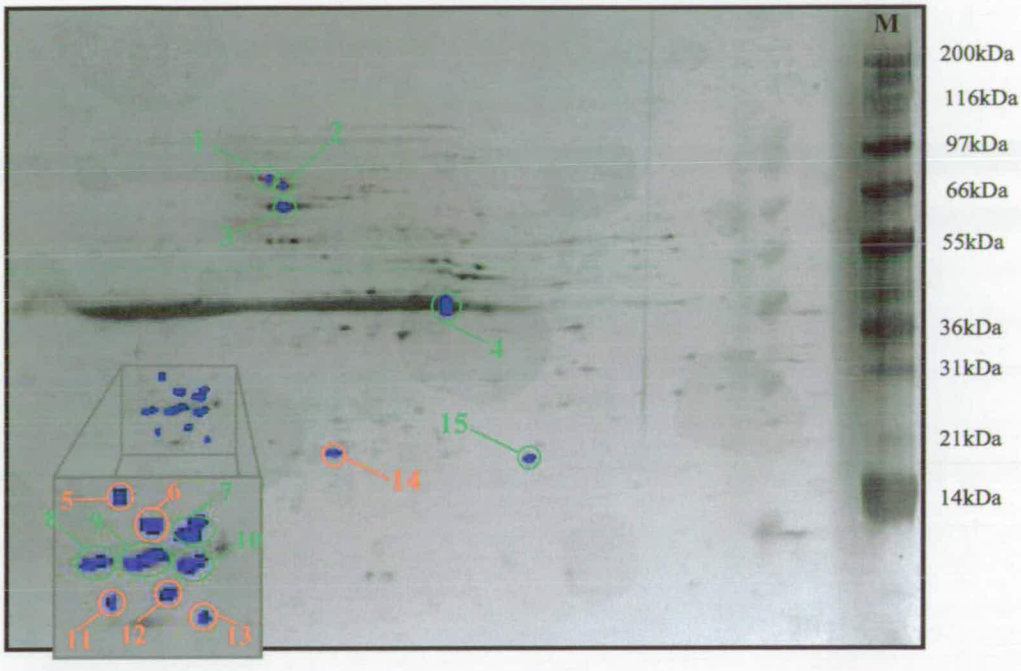


Figure 3.35 – Spots excised (shown **blue**) for subsequent in-gel digestion with trypsin and MALDI-ToF-MS analysis. Identified and non identified spots labelled in **light green** and **red** respectively. Labelled regions referred to in discussion. **M** – molecular weight marker (Mark12, Invitrogen) [200kDa = Myosin, 116kDa = α - galactoside, 97kDa = Phosphorylase b, 66kDa = BSA, 55kDa = Glutamic dehydrogenase, 36kDa = Lactate dehydrogenase, 31kDa = Carbonic anhydrase, 21kDa = Trypsin inhibitor, 14kDa = Lysosyme.]

($\approx 4x10^{+11}$) for the identity of these species both being Heat Shock Protein 70 (Hsp70, Ac# K01298) which suggests very low probability of random identification. Hsp70 has a mass of approximately 69kDa and pI of 4.83 which both correspond to the mobility of this species we observe in this 2DE experiment adding further to the evidence of the correct identification. MS fingerprint data of species 3 shows even greater statistical probability for its top hit which in this case was GroEL with a score of $1.7x10^{+17}$, the GroEL protein has mass and pI of 57kDa and 4.82 respectively which again shows good consistency with the mobility of the species 3 we observed. During the gel comparisons we identified that biotin synthase shows a huge amount of streaking across the pH range

between 4 and 5 which is again seen in this gel however the streaking is more pronounced but the section of this streak shown by section 4 in Figure 3.35 was analysed alongside the unknown species and this surprisingly produced a lower statistical score than the two previous samples of $2.63x10^{+6}$ but this remains good conformation of the identity of biotin synthase. The lower statistical score could be directly due to the streaking which could cause a decrease in protein concentration in a specific area. the theoretical mass and pI make this an unlikely candidate being the region is due to the degree of loading it was impossible to identify other species that may lie within this region but these could affect the resultant mass list and subsequent database search. Gel portion 7 shows a very strong statistical hit for a putative sidephore receptor (IreA) but was excised from a region approximately 30kDa and pH 4.25 suggesting an incorrect identification and may demonstrate the importance of initial arbitrary information from the 2 Dimensional Gel. The other hit for species 7 and hits for species 8, 9 and 10 are *E. coli* biotin synthase with statistical probabilities varying from 1890 down to 375. Although these hits are of fairly low probability upon further analysis it was clear that the peptides identified corresponding to biotin synthase were located within fairly restricted regions of the linear sequence. These species, are therefore assumed to be major breakdown products of biotin synthase and not a separate species up-regulated as hoped for. The final species identified was an outer membrane protein found from the spectra obtained from region 15 of the gel. This showed high statistical probability and theoretical mass and pI correlation enhancing the identification.

3.7.6 – Conclusions of Differential Expression Analysis of Biotin Synthase

Our aim form the differential analysis was to identify possible *in vivo* protein partners of biotin synthase by means of up-regulation due to the over-expression. As discussed above the protein spots initially identified as the main targets of the analysis were eventually identified to be breakdown products of biotin synthase. Several species were identified easily from the gel proving the validity of the experimental conditions.

Unfortunately in this analysis, we were unable to demonstrate that the over-expression of biotin synthase results in the appearance or upregulation of cell protein component(s).

Gel portion	Protein ID	Species	Mass(kDa)/pI	Accession #	Mouse Score
1	Hsp70	*E. coli*	69 / 4.83	K01298	$4.7x10^{+11}$
2	Hsp70	*E. coli*	69 / 4.83	K01298	$3.6x10^{+10}$
3	GroEL	*E. coli*	57 / 4.82	AE000487	$1.69x10^{+17}$
	Unknown	*E. coli*	58 / 4.85	X07850	$1.7x10^{+17}$
4	Biotin synthase	*E. coli*	38.5 / 5.32	AE000180	$2.63x10^{+6}$
7	Putative sidephore receptor (IreA)	*E. coli*	75 / 6.04	AF320691	$1.34x10^{+3}$
	Biotin synthase	*E. coli*	38.5 / 5.32	AE000180	658
8	Biotin synthase	*E. coli*	38.5 / 5.32	AE000180	513
9	Biotin synthase	*E. coli*	38.5 / 5.32	AE000180	$1.89x10^{+3}$
10	Biotin synthase	*E. coli*	38.5 / 5.32	AE000180	375
15	Outer Membrane protein	*E. coli*	27 / 9.16	L37088	$9.77x10^{+4}$

Figure 3.36 – Table showing database hits for fingerprint spectra found for species highlighted in 2D-Gel in Figure 3.35. Species where no strong hits were found are excluded.

3.8 – Structural Analysis of purified 6His-Biotin Synthase

3.8.1 – Anion Exchange Chromatography of purified Biotin Synthase

Several distinct species of expressed protein were observed during the 2 Dimensional Gel Electrophoresis analysis of expressed biotin synthase. Thus we concluded that several distinct and definable ionic species of the protein exist. These different species are normally unrecognised due to co-elution of these during low resolution anion exchange, Ni affinity chromatography and standard 1D SDS-PAGE. To elucidate and differentiate these species we have performed high resolution analytical anion exchange chromatography on the purified 6His-BS protein using the Amersham-Biosciences 1 ml MonoQ column. This column has a maximum binding capacity of 10 mg and is suitable for use with a large range of compounds, including protein mixtures.

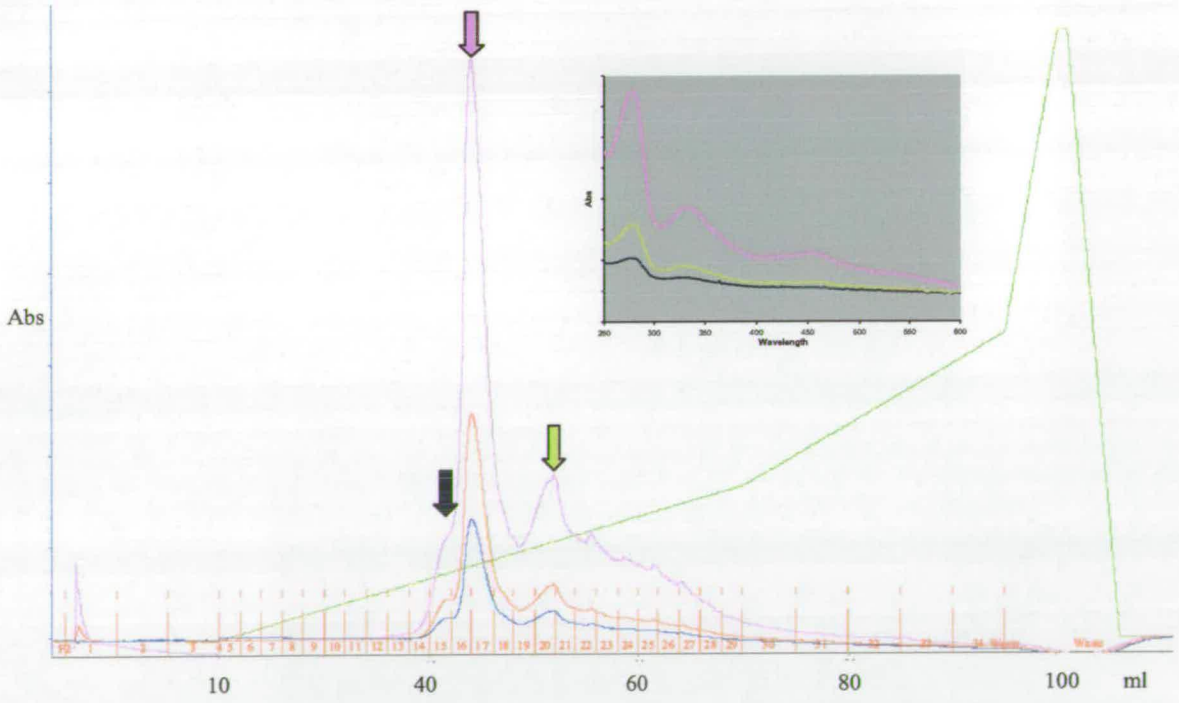


Figure 3.37 – Elution profile of 6His-BS from MonoQ anion exchange column, during a 0 to 500mM NaCl gradient. Monitored wavelengths are **280nm**, **330nm** and **425nm** are shown and 3 major species indicated with arrows. Inset is UV-Vis 250-600nm scan of three major species in there corresponding colours.

During this analysis approximately 10 mg of purified biotin synthase (as judged by Bradford Assay (166)) which was stored in 50 mM Tris-HCl, pH7.5, the sample was loaded at 1 ml/min. During loading the top of the column packing was visually red in colour. The mobile phases used were Buffer C and D as described in Section 2.2 and the elution profile was optimised using 6His-BS. A 0 to 500 mM NaCl gradient over 100 CV was used to elute protein from the column and elution was monitored at 280 nm, 330 nm and 425 nm wavelengths, 1 ml fractions collected. Column was cleaned with a 500 mM to 2 M NaCl gradient and re-equilibrated with 0 mM NaCl for 30 min.

We identified several species in our purified 6His-BS and these produced a complex elution profile with the major peaks eluting at approximately 218 mM, 240 mM and 314 mM NaCl. This profile from this high resolution analysis differs to the previously observed single peak for other chromatography profiles. A plot of the elution profile for 6His-BS is shown if Figure 3.37 and a similar profile is observed for all the monitored wavelengths.

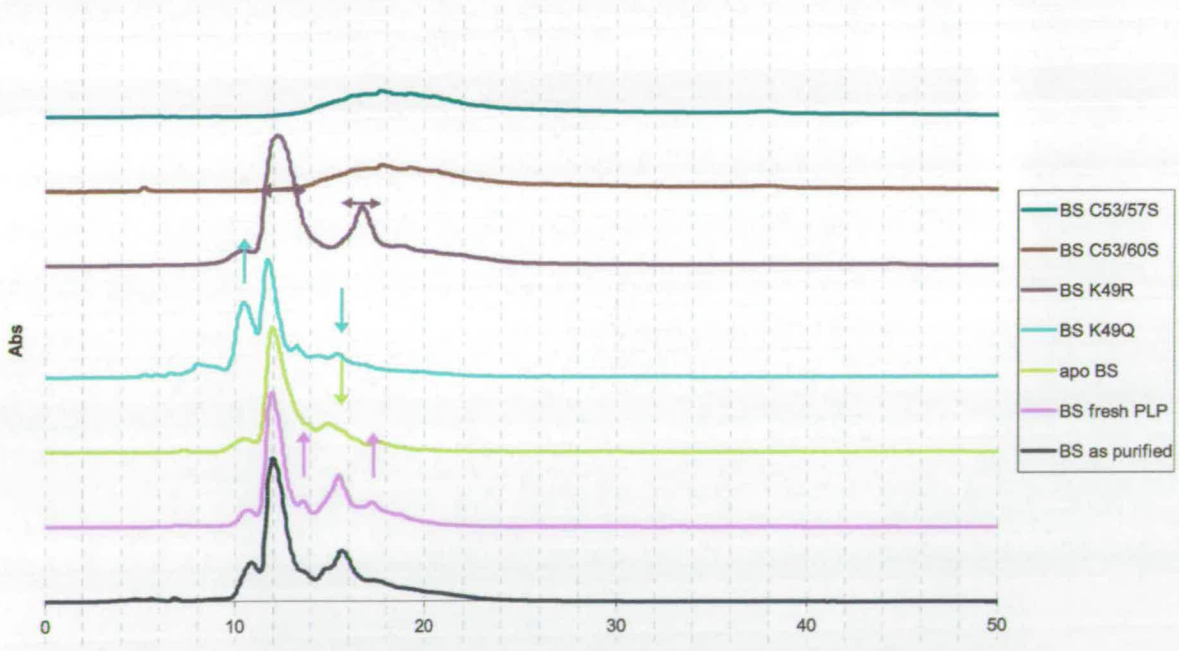


Figure 3.38 – Pileup of elution profiles for 6His-BS, apoBS fresh PLP loaded BS and 6His-BS muants. Trace shows absorbance at 280nm of sample across 0-500mM NaCl gradient. Profiles are offset by 200units to clarify diagram. Differences from 6His-BS are indicated.

The wavelengths 330 nm and 425 nm were monitored in addition to 280 nm, as these are typical maxima for [Fe-S] cluster containing species. These observations suggest that all species contain some form of [Fe-S] cluster and that the observed phenomenon is not the result of holo and apo forms of the protein. Inset in Figure 3.37 is the progressive UV-Vis scan of the 3 major peaks (1, 2 and 3), indicated with arrows. These all show identical absorbance characteristics which importantly display the maxima observed for the purified BS as described in Section 3.3.

The experiment was repeated using the same conditions as described above with apo-BS BS freshly loaded with PLP and all biotin synthase mutants available to us as described in Section 3.2. The elution profiles for all BS mutants, 6His-BS and apo-BS are shown in Figure 3.38, this figure shows the absorbance measured at 280 nm only, for clarity and variations from 6His-BS are highlighted. All differences observed are reproduced in all 3 monitored wavelengths (280 nm, 330 nm and 425 nm). Biotin synthase freshly loaded with PLP shows little difference, however we do observe a small increase in the absorbance of the satellite peaks observed adjacent to the major peaks 2 and 3. The trace for apoBS shows a complete lack major species 3 but does display the other peaks, including the satellite peaks. The observed satellite peaks seem to be enhanced by loading fresh PLP. The BS-K49Q mutant demonstrates an enhancement of the first elution species to equivalent absorbance value of the major species observed for BS wild-type, however similarly to apoBS this protein shows an almost complete lack of the species 3. In comparison the BS-K49R mutant resembles the 6His-BS trace to a large degree however much peak broadening is observed for all 3 major peaks. Clearly the greatest difference in profiles was observed for the double cysteine mutants, C53/60S and C53/57S, which both display a complete lack the definition which is observed for the other protein samples. Both double mutants show a single broad peak which elutes with a front edge at approximately 250 mM NaCl, peaking at around 370 mM NaCl and trailing edge above the 500 mM limit.

3.8.2 – Size Exclusion Chromatography of Purified Biotin Synthase

The analysis of the various biotin synthase proteins was also performed using the Superdex200 analytical, 36ml, Size Exclusion column from Amersham-Biosciences. The Superdex200 column has a dynamic separation range from 10kDa to 600kDa which would allow separation of monomer and up to and including a BS octomer.

The protein samples were loaded, approximately 1 mg / sample / run, and a 40ml isocratic elution performed at 1 ml/min using 50mM Tris-HCl, pH7.5 + 100mM NaCl and elution monitored at 280, 330 and 425nm. As shown in Figure 3.40 BS wild-type showed similarly complex profile but displays this with apparently inverted topography to that observed for anion exchange analysis. The lowest abundant peak is observed first during the elution and this displays a distinct shoulder with lower mobility under these conditions.

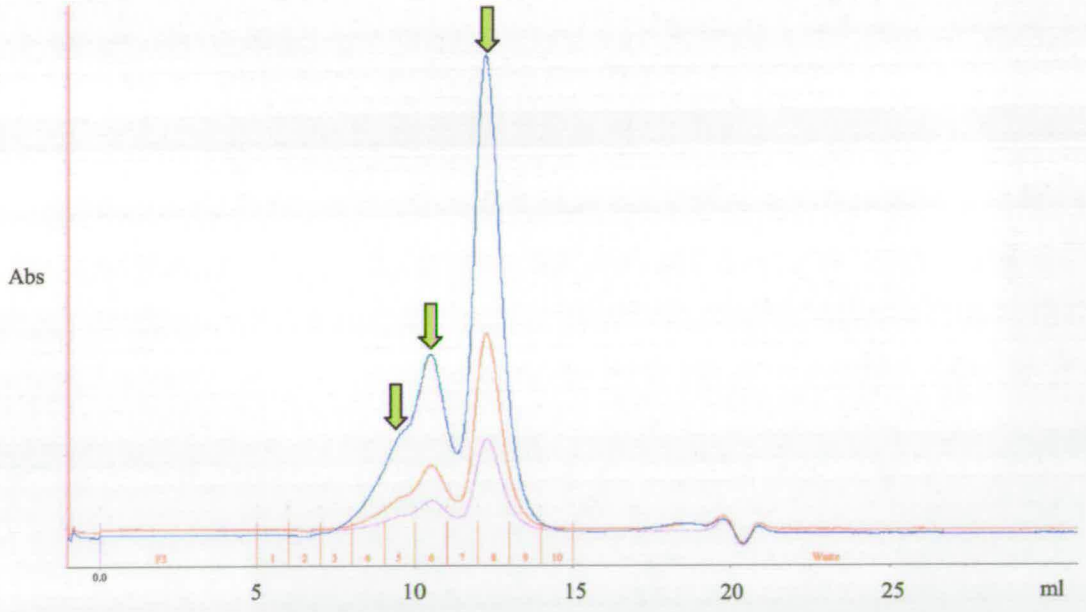


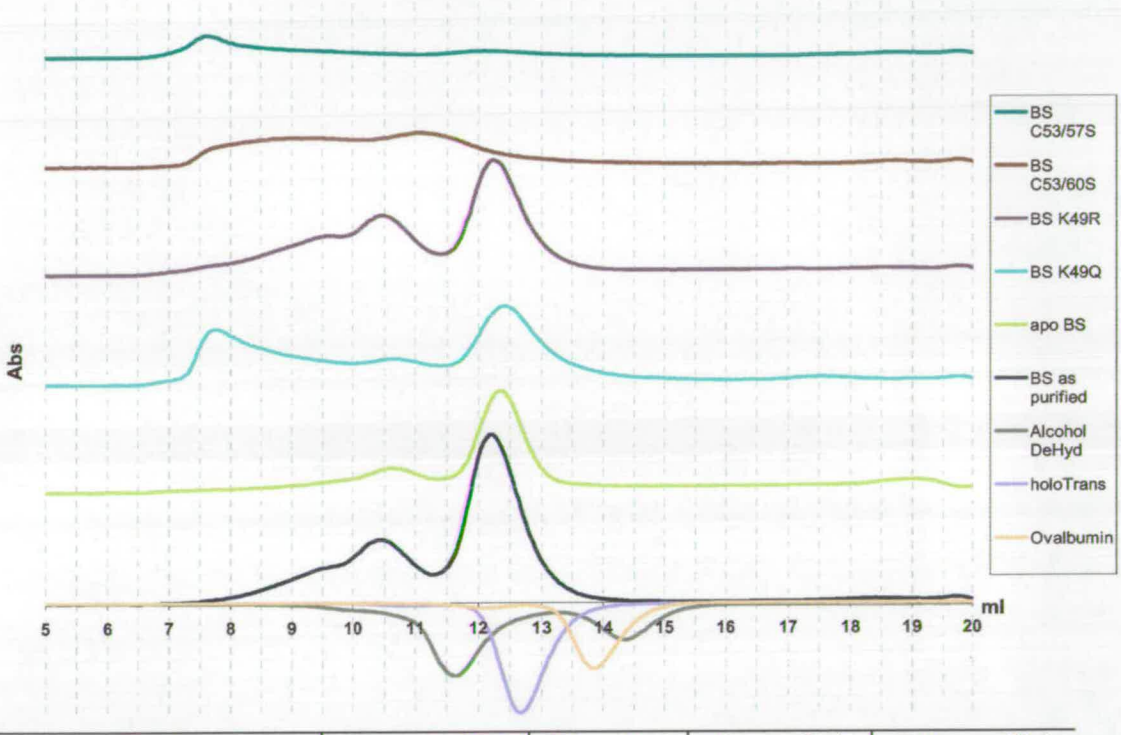
Figure 3.40 – Elution profile of 6His-BS from Superdex 200 size exclusion column, during a 36ml isocratic elution with 50mM Tris-HCl, pH7.5 + 100mM NaCl. Monitored wavelengths are **280nm**, **330nm** and **425nm** are shown and 3 major species indicated with arrows.

The major species follows and this demonstrates a very sharp elution returning to baseline within 1-2 ml. As observed in the anion exchange analysis all of the observed species demonstrate absorbance at 280, 330 and 425 nm suggesting all contain some form of [Fe-S] cluster. As these species demonstrate similar profiles for all monitored wavelengths this comparison we focus on the 280 nm wavelength for clarity. Again large differences are observed between the BS mutants, 6His-BS and apoBS. The profile of apoBS is very similar to that of 6His-BS but absorbance at 330 nm and 425 nm are almost totally absent. The main difference lies in the lack of the shoulder present prior to the first elution species as indicated in Figure 3.41. The profile of the BS-K49Q protein displays significant differences to the 6His-BS in that a formation of a highly abundant species of higher mobility, however it the two major species observed for 6His-BS are still present but with lower relative abundance. Of all the mutant proteins the profile of BS-K49R resembles 6His-BS to the largest degree and the only observable difference is the resolution of the early eluting shoulder species into a peak, overlapping with the later higher abundant peak. Both double cysteine mutants show a long shallow elution profile from the high molecular weight range and this continues to around the 100kDa range. BS C53/57S demonstrates a more pronounced higher mobility species similar to that seen for BS-K49R but BS-C53/60S seems to be an even distribution across the range.

The molecular weight of the observed species can be gauged from the elution of the three standard proteins, Alcohol Dehydrogenase (150kDa), holoTransferrin (81kDa) and Ovalbumin (43kDa) and their theoretical masses, inset Figure 3.41.

We observed no monomeric BS species based on the monomeric 43kDa Ovalbumin standard protein. The major species, in the case of 6His-BS, displays an elution time preceding the 81kDa holoTransferrin standard and after the 158kDa Alcohol Dehydrogenase standard. This elution time would suggest a wide dimer configuration of BS. Previous experiments with native gel electrophoresis suggest predominantly a dimeric structure of 104kDa with a [2Fe-2S] cluster per monomer (68). A study by Hernandez *et al.* utilises native protein Mass Spectrometry to observe a variety of monomer and dimer species containing a variety of [Fe-S] clusters and other modifications (173). With these studies in mind our observations make it likely that the

elution times correspond to an 'open conformation' dimer with bound [Fe-S] cluster(s). The second major species, displaying higher mobility, could perhaps correspond to a tetramer with similar modifications. As described above the BS-K49Q and double cysteine mutant proteins show a variety of species of much higher apparent MW which could be due to the formation of a large multimeric compexes or chains with little 3 dimensional structure which thread through the column packing.



Protein	Monomer (Da)	Dimer (Da)	Trimer (Da)	Tetramer (Da)
6HisBS / BS-K49Q / apoBS	39,670	79,340	119,010	158,680
BS-K49R	39,698	79,396	119,094	158,792
BS-C53/57S / BS-C53/60S	39,638	79,276	118,914	158,552

Figure 3.41 – Pileup of isocratic elution of biotin synthase and mutant proteins monitored at 280nm. Standard protein traces are inverted (Alcohol Dehydrogenase (150kDa), holoTransferrin (81kDa) and Ovalbumin (43kDa). Theoretical apo-BS quaternary structure molecular weights are shown in table.

3.8.5 – Limited Digestion of as Purified and Prepared Apo Biotin Synthase Protein Samples

Aerobically prepared BS in reported to contain one or two [2Fe-2S] (68, 90, 7, 174). To investigate the effect of the presence of this cluster on protein structure a limited digestion experiment was performed. Samples of our purified 6His-BS and apoBS were subjected to a limited digestion experiment. The apoBS was prepared as described in Section 2.20 and as previously reported (164, 141, 86) anaerobically in N_2 with sodium dithionite and EDTA and apoBS was re-purified from reduction buffer with PD10 gel filtration column before removal to normal atmospheric conditions.

Protein samples (1 mg) were incubated with 5 ng of the endoproteinase Asp-N for specific time periods and reactions stopped with addition of SDS-sample buffer and boiling for 5min. All samples were run on SDS-PAGE gel as shown in Figure 3.42 and protein bands were visualized using GelCode reagent (Pierce). The band intensity can be used as a comparative protein concentration indicator with a good linear dynamic range as equal quantity was loaded in each lane. The gel was analysed with the Phoretics 2D software (NonLinear) to show band intensity along a single SDS-PAGE gel lane. Time course profiles for each sample are shown either side of gel and traces are shown in corresponding colours as lane label. The BS wild-type time course shows clear formation of several species of lower molecular weight than the monomer within 10 min and the intensity of these species remains fairly constant throughout the 25min time period. The species observed at higher molecular weight are thought to be protein dimer and this is slowly reduced over the 25min incubation. The intensity of the "parent" monomer band appears to remain constant but as concentration of protein used was 10 μg/μl it was in great excess and most likely above the upper limit of the protein stain. Clearly species with similar molecular weight are formed in the apoBS digestion however they are formed in a shorter time period and to much higher relative concentration. This suggests a greater lability of the protein to AspN proteinase digestion. However no species exclusive the apoBS are observed. Although the [Fe-S] cluster appears not to directly block an accessible AspN site, the structure is considerably more open or disordered allowing a much more rapid digestion of the

protein. In both cases we observe 6 major bands with apparent MW greater than 6 kDa which could correspond to a single cleavage at solvent exposed Asp residue. Subsequent cleavages resulted in the digestion of the protein into smaller species. To help in the characterisation of these species the gel was calibrated using the using the Phoretics 2D software and themolecular weight marker as standards. This allowed estimations of the molecular weight of each observed protein bands to be made. Figure 3.43 shows a single lane from the gel above and the molecular weight marker, the associated table shows the calibrated molecular weight for each species based on a First Order La Grange migration of the marker proteins. Inevitably this type of MW estimation of gel migrated species is inherently inaccurate and we believe that species A and B are likely to be the species observed by Sanyal *et al.* by native gel electrophoresis of 82kDa and 104kDa. These are attributed to homodimers with one [2Fe-2S] per dimer and two [2Fe-2S] per dimer respectively. The observation of these species under the denaturing and reducing conditions of SDS-PAGE suggests very tight complexation and binding of the clusters. The estimation of band C, presumed biotin synthase monomer shows good correlation with the calculated 39,670 Da molecular weight of the 6His-BS linear protein sequence. Observed species of calibrated molecular weight range from approximately 29 kDa to 8 kDa and are the proteolytic fragments of biotin synthase as described above.

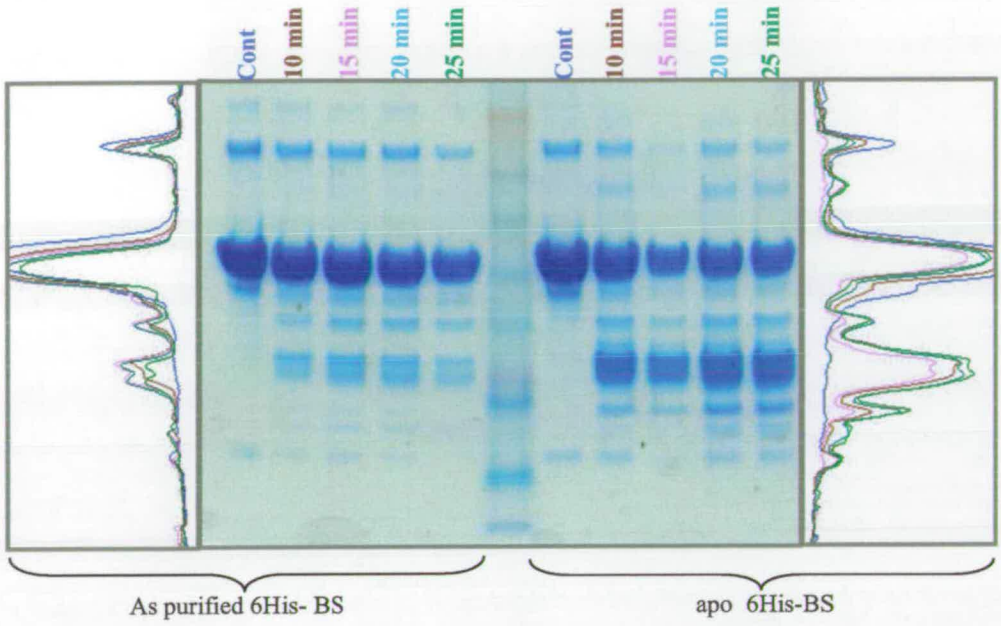


Figure 3.42 – Profile analysis of Limited digestion of as purified and apo biotin synthase with endoproteinase Asp-N. Traces at side of gel are intensity plots for each lane in corresponding colour. Centre lane is SeeBlue2 molecular weight marker. [188kDa = Myosin, 98kDa = Phosphorylase b, 62kDa = BSA, 49kDa = Glutamic dehydrogenase, 38kDa = alcohol dehydrogenase, 28kDa = Carbonic anhydrase, 18kDa = Myoglobin, 14kDa = Lysosyme, 6kDa = Aprotinin, 3,5kDa = unresolved Insulin]

Band ID	Calibrated MW (kDa)
A	104.7
B	78.2
C	39.6
D	35.7
E	29.1
F	21.6
G	18.3
H	13.1
I	11.4
J	8.4

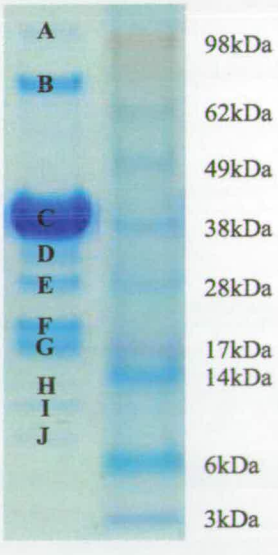
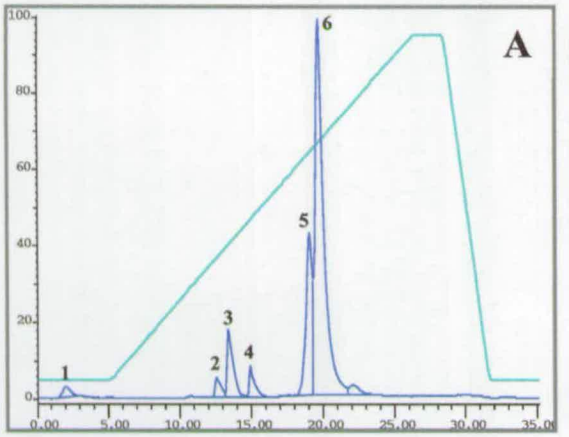
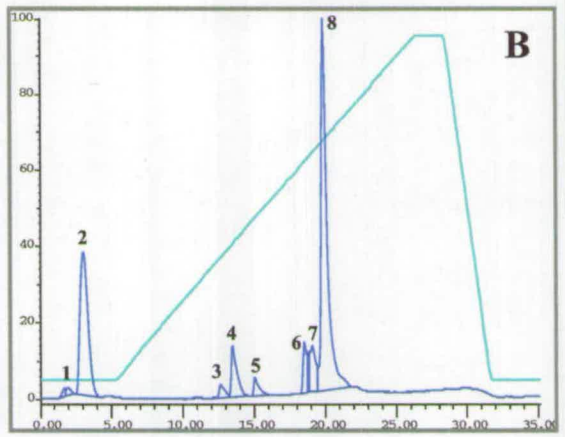


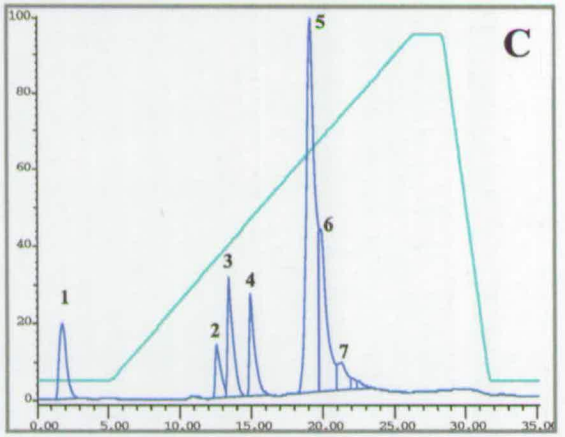
Figure 3.43 – Calibrated molecular mass of species observed during the limited digestion of as purified biotin synthase protein with Asp-N. Calibration performed using the SeeBlue2 molecular weight marker as mass standards. [188kDa = Myosin, 98kDa = Phosphorylase b, 62kDa = BSA, 49kDa = Glutamic dehydrogenase, 38kDa = alcohol dehydrogenase, 28kDa = Carbonic anhydrase, 18kDa = Myoglobin, 14kDa = Lysosyme, 6kDa = Aprotinin, 3,5kDa = unresolved Insulin]

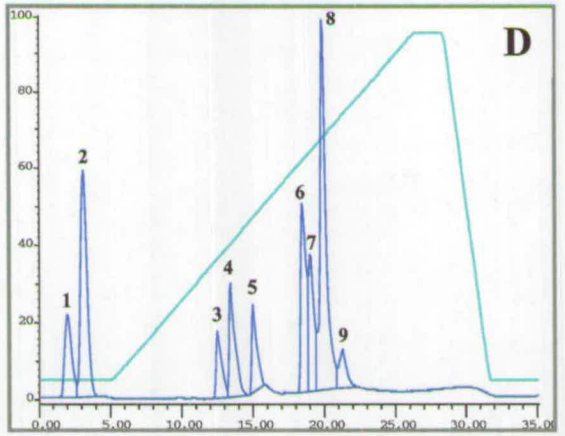
3.8.6 – HPLC-RPC of Limited Digestion samples of Biotin Synthase

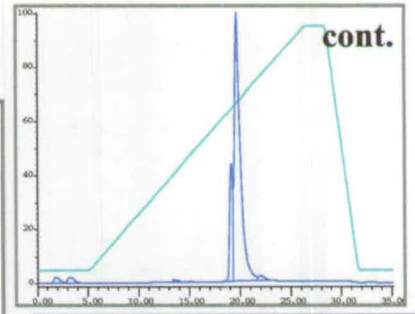
For further analysis and purification of samples for mass spectrometry the limited digestions were repeated on large scale, small quantity of sample removed for SDS-PAGE and remaining sample incubated with or without DTT for 15min at room temperature. These samples were directly loaded on C5-RP-HPLC column. The HPLC elutions were monitored at 280 nm and 1 min fractions collected throughout the elution. As observable in Figure 3.44 the elution profiles differed greatly between apoBS and as purified 6His-BS as previously discussed. Interestingly differences were also observed between reduced and non reduced samples. All fractions were collected and stored at -20°C before further analysis. The reduction of the samples yields peaks just preceding the major species, undigested 6His-BS, the enhancement of these may due to the formation of enzymatic fragments, which remain linked to the main species until reduction. The 3 peaks of retention time in the region of 13 to 16 min show increased intensity for apoBS compared to as purified which suggests the intensity is similarly affected by the presence of the [Fe-S] cluster as observed for some of the species in Section 3.8.5 and may indeed be the protyolytic fragments observed on the SDS-PAGE gel.











Peak	Retention time (min)			
	A	B	C	D
1	1.93	1.96	1.67	1.93
2	12.46	2.97	12.50	3.04
3	13.40	12.57	13.34	12.43
4	14.87	13.39	14.91	13.35
5	19.00	14.98	19.10	14.95
6	19.60	18.47	19.84	18.43
7		19.06	21.25	18.99
8		19.84		19.81
9				21.25

Figure 3.44 – Traces of 280nm monitored C5-RP-HPLC purification of limited digestion of biotin synthase samples. A = 6His-BS, B = 6His-BS reduced, C = apo6His-BS, D = apo6His-BS reduced. Control sample trace is shown and all samples shown as run on SDS-PAGE with molecular weight marker (Sigma, 96kDa, 66kDa, 45kDa, 30kDa, 26kDa and 14kDa)

3.8.7 – Mass Spectrometry of HPLC Purified Limited Digestion Fragments of Biotin Synthase

Each fraction from the HPLC purification described in Section 3.8.6 was spotted with Sinapinic acid Matrix directly onto MALDI target plate and analysed using the Applied Biosystems Voyager DE™STR using a low and high lazer power to maximise observation of peptides from all mass ranges. Spectra were acquired for all fractions in samples A, B, C and D from Figure 4.44, species observed above 6 kDa were used in the subsequent analyses. Spectra were analysed and each mass list used to search theoretical Asp-N digestion allowing identification of fragments in each fraction.

Exclusion of peptides less than 6kDa was to filter the mass list the range observable by SDS-PAGE. The masses of observed species meeting these criteria were assembled and duplicates removed providing definitive list of observed protein/peptide species found for each sample (A to D). The final mass list was used to search the theoretical biotin synthase Asp-N map as previously described using PAWS software the results of which are shown in Figures 3.45 to 3.48. The Figures show all obtained masses and their identified section of the BS sequence identified according to measured mass. Identified regions are referenced and coverage map shows region of protein found for each species. Two abundant species are seen at mass 38,445Da and 38,662Da which correspond to the final 11 and 9 residues being cleaved from the protein resulting in a species which would be virtually indistinguishable from undigested protein by SDS-PAGE Ref # 5 and 6. Fragments of the protein are observed originating from a single cleavage at leaving the N-term intact (Ref # 1,2,3 and 7) suggesting the Asp sites 64, 117, 156, 189 and 242 are solvent exposed and available for digestion with the protease when the 6His-BS is in buffered solutions.

[A] – 6His-BS non reduced		
Ref. #	**Mass Obs. (Da)**	**Match**
1	7,450.52	M[1-63]E
2	13,454.38	M[1-116]R
3	17,593.20	M[1-155]L
4	26,975.95	M[1-241]A
5	38,445.97	M[1-343]P
6	38,662.16	M[1-345]T
7	21,665.88	M[1-189]R
8	9,894.25	D[30-116]R
9	8,305.35	D[117-189]R
10	14,431.43	D[117-249]F
11	9,400.76	D[156-241]A
12	10,292.61	D[156-249]F
13	8,480.82	D[163-241]A
14	13,373.50	D[183-306]E
15	7,540.56	D[242-308]K
16	11,704.230	D[242-345]T
17	12,486.09	D[244-354]L
18	10,812.38	D[250-345]T

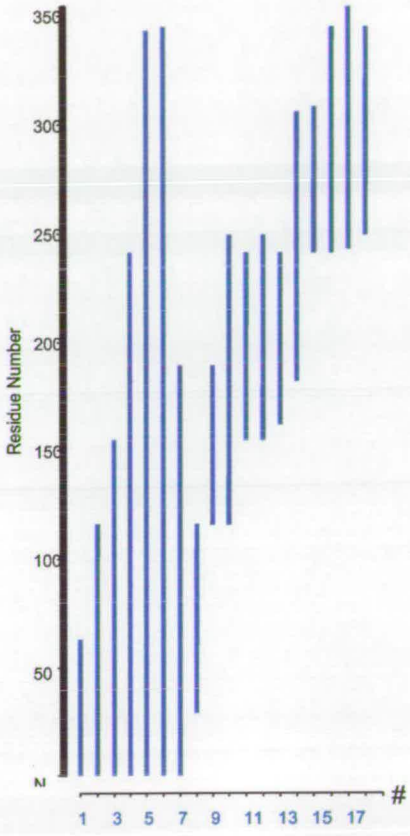


Figure 3.45 – Table and PAWS map of peptides of mass >6kDa found for limited digest of as purifiec 6His-BS after purification on HPLC and MALDI-ToF and PAWS analysis.

A significant number of the observed species seem to originate or terminate within the region of D242. All species containing peptides after this region originate from this narrow area of the protein sequence. With the exception of the almost entirely undigested protein only one species was observed that traversed this D242 region (Ref # **14)** which suggests extremely high lability of this site/region for proteolysis and a therefore a possibly disordered linker/loop region. The reduced sample of 6His-BS digestion (B – Figure 3.46) displays many similarities with the non reduced sample including the ladder of species consisting of residue 1 to 63, 116, 155, 162, 182 and 241 (Ref # 1-6). However it does contain more species whose mass would correspond to species with 2 Asp sites cleaved. These are tend to be toward the N-terminal end of sequence.

This results in species such as D[44-116]R shown as Ref # 8, 9 and 10. In this case 4 species are observed that traverse the D242 region (Ref# 16, 17, 18 and 19) which all suggests that the reduction of the protein with DTT has somehow destabilised the protein structure, and as such some further Asp sites are made available.

The two samples of digested apoBS show very similar coverage and fragments to the reduced as purified 6His-BS as shown in Figures 3.47 and 3.48 but again it is clear that more species are observed when the sample is reduced after the limited digestion. Interestingly in the case of both reduced samples the final 26 amino acids are not observed whereas the whole sequence is covered when the digested sample is left unreduced. This may suggest that the C-term of the protein is linked via a reducible chemical bond which would be typically a disulfide. Although less pronounced, the region containing residues 242 to 250 seems to be a labile region and the existence of 5 Asp residues within this, results in a number of cleavages that yield a singly dissected protein chain.

[B] – 6His-BS reduced

Ref. #	Mass Obs. (Da)	Match
1	7,452.52	M[1-63]E
2	13,454.38	M[1-116]R
3	17,593.20	M[1-155]L
4	18,513.15	M[1-162]L
5	20,899.75	M[1-182]L
6	26,975.95	M[1-241]A
7	38,445.97	M[1-343]P
8	9,894.25	D[30-116]R
9	8,144.31	D[44-116]R
10	10,158.69	D[64-155]L
11	14,307.20	D[64-189]R
12	8,305.35	D[117-189]R
13	13,539.59	D[117-241]A
14	9,629.96	D[156-243]N
15	16,680.04	D[156-306]E
16	16,938.47	D[156-308]K
17	12,531.54	D[190-306]E
18	10,746.45	D[209-306]E
19	7,540.56	D[242-308]K
20	11,704.23	D[242-345]T
21	8,813.24	D[250-328]G

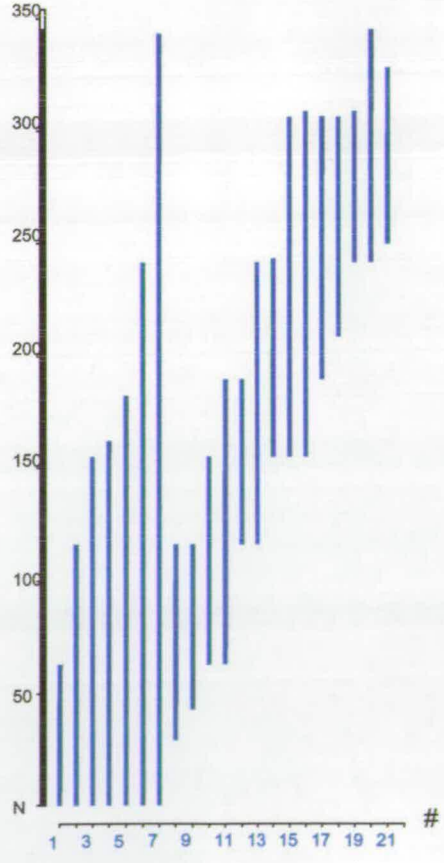


Figure 3.46 – Table and PAWS map of peptides of mass >6kDa found for limited digest of reduced as purified 6His-BS after purification on HPLC and MALDI-ToF and PAWS analysis.

This interesting data provides some convincing evidence as to some of the solvent exposed regions of the protein. However as the data is based on mass only identification and a further study utilising duel digestion was undertaken to produce some more convincing evidence that the observed masses have been correctly assigned to the protein chain.

[C] – apo6His-BS non reduced		
Ref. #	**Mass Obs. (Da)**	**Match**
1	13,454.38	M[1-116]R
2	17,593.2	M[1-155]L
3	18,513.15	M[1-162]L
4	20,899.75	M[1-182]L
5	26,975.95	M[1-241]A
6	38,662.16	M[1-345]T
7	9,894.24	D[30-116]R
8	10,158.69	D[64-155]L
9	27,063.98	D[64-308]K
10	7,463.39	M[1-63]E
11	10,090.44	D[117-208]K
12	13,539.59	D[117-241]A
13	9,745.04	D[156-306]D
14	16,680.04	D[156-306]E
15	6,094.22	D[183-241]A
16	12,531.54	D[190-306]E
17	7,540.56	D[242-308]K
18	11,704.23	D[242-345]T
19	12,715.28	D[242-354]L
20	6,648.71	D[250-308]K
21	10,812.38	D[250-345]T

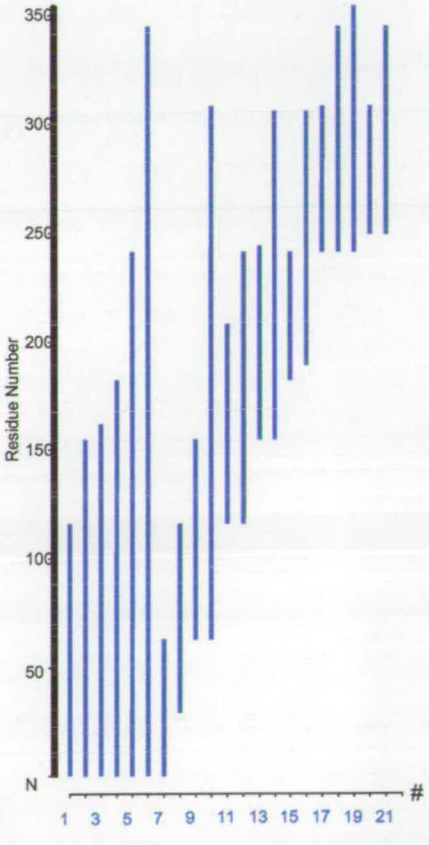


Figure 3.47 – Table and PAWS map of peptides of mass >6kDa found for limited digest of reduced as purified apo-BS after purification on HPLC and MALDI-ToF and PAWS analysis.

[D] – 6His-BS reduced		
Ref. #	**Mass Obs. (Da)**	**Match**
1	13,454.38	M[1-63]E
2	26,975.95	M[1-241]A
3	17,593.2	M[1-155]L
4	18,513.15	M[1-162]L
5	20,899.75	M[1-182]L
6	7,452.52	M[1-63]E
7	23,974.33	D[30-246]V
8	6,019.87	D[64-116]R
9	19,541.44	D[64-241]A
10	11,078.64	D[64-162]L
11	27,063.98	D[64-308]K
12	14,431.43	D[117-249]F
13	10,090.44	D[117-208]K
14	16,680.04	D[156-306]E
15	8,480.82	D[163-241]A
16	6,094.22	D[183-241]A
17	12,531.54	D[190-306]E
18	10,746.45	D[209-306]E
19	13,154.24	D[209-328]G
20	11,704.23	D[242-345]T
21	6,738.79	D[247-306]E
22	10,812.38	D[250-345]T
23	8,813.24	D[250-328]G

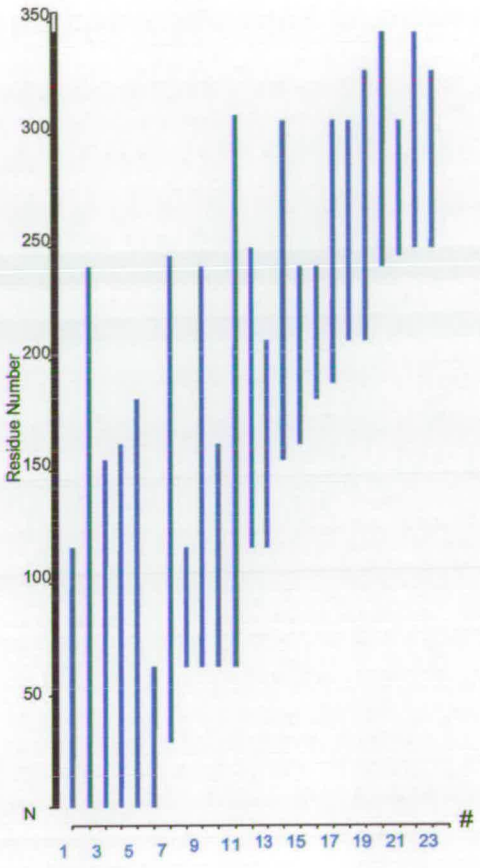


Figure 3.48 – Table and PAWS map of peptides of mass >6kDa found for limited digest of reduced as purified apo-BS after purification on HPLC and MALDI-ToF and PAWS analysis.

3.8.8 – Dual digestion of Biotin Synthase for Structural Mapping

To further substantiate that the limited digestion fragments of biotin synthase identified in Section 3.8.7 were correctly assigned a scheme was devised involving a dual endoproteinase digestion. This samples were analysed using MALDI-ToF and data compared using the PAWS maps of the protein. The scheme is outlined in Figure 3.49. Essentially the initial limited digestion was performed and SDS-PAGE used as a 1D separation technique.

Bands were observed using coomassie stain (as shown in Figure 3.42/43). Bands were excised and in-gel digestion with trypsin performed on each. The doubly digested samples were prepared as previously described for MALDI-ToF and acquired peptide spectra deisotoped. The resultant mass lists were used to search the PAWS trypsin digestion map. The identified peptides were then super imposed onto the theoretical PAWS Asp-N digestion map. The observation of a trypsin peak from a specific Asp-N fragment was used as evidence of the presence of that Asp-N derived peptide in the band excised (see figure 3.49).

Bands C and E to J (Figure 3.43) were excised from the gel and digested with trypsin as described above. The result of this analysis is shown in Figure 3.50 where identified Asp-N regions are shown compared to linear sequence. The C_{term} of the linear sequence was not observed at any point and this could be due to these peptides not ionising or this region is very quickly digested. If this is the case species B and C, are the identified species Ref 5/6 from Figure 3.44. In this case the intact protein has had final 11 and 9 amino acids cleaved respectively. Similarly no evidence was found for the presence of Asp-N fragments 6, 7 and 8 in Band C, or any other bands, and this could again be due to a double cleavage causing a extraction of a region of the chain from the protein, however no species corresponding to this mass was observed. Band E shows a coverage matching the essentially intact protein up to the Asp rich region 242-249 exclusively and no evidence is found for any peptides toward the $C_{term.}$

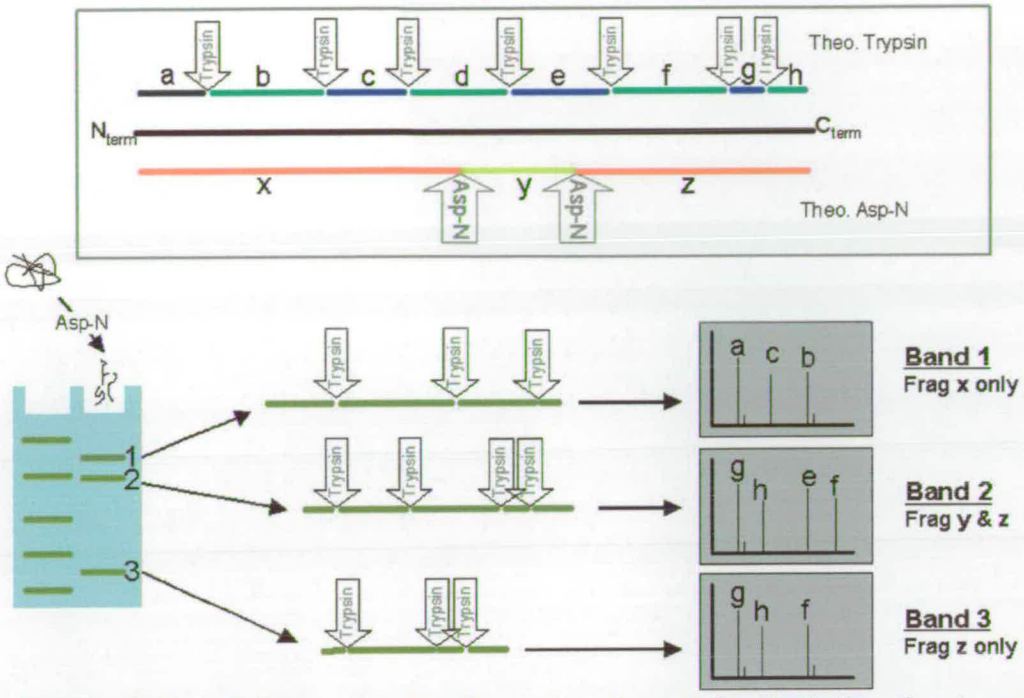


Figure 3.49 – Scheme showing the protocol used for dual digestion to provide further evidence for previously identified Asp-N fragments of biotin synthase. An initial limited digestion was separated by SDS-PAGE and bands individually excised and in-gel digestion was performed with trypsin. Trypsin fragments were identified with MALDI-ToF and compared to theoretical Asp-N digestion map, thus showing existence of a specific Asp-N derived peptide in corresponding gel species.

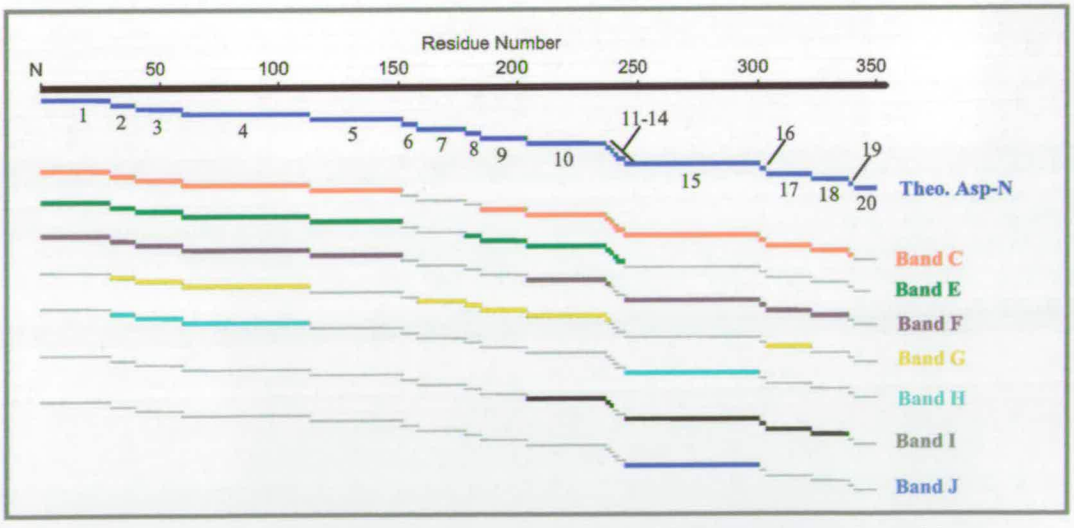


Figure 3.50 – Map of Asp-N derived peptides of biotin synthase found in bands C and E to J using dual digestion method. Species are shown offset and in corresponding colours compared to linear sequence and theoretical digestion.

As suggested (shown in Figure 3.43, Ref 4) this region, is cleaved quickly and efficiently, and thus resulting in two sections of biotin synthase. The C_{term} portion of this single cleavage seems to be very labile compared to the N_{term} as no evidence is found for this C_{term} fragment intact. However this is strong evidence that the section to the N_{term} side of this region is both soluble and resistant to digestion and so structurally distinct. The hypothesis of this region being a boundary between domains is strengthened by this evidence. The remaining bands produced a variety of identified regions of the protein which correspond to species with very similar molecular weight to others and so would be found with similar SDS-PAGE mobility. Several species identified in Figure 3.43 can be paired with the equivalent region of the protein identified with the dual digestion methodology. As discussed above Band C shows constancy with species Ref 5 and 6 also Band E demonstrates good consistency with species Ref 4. All consistent data is summarized in Figure 3.51.

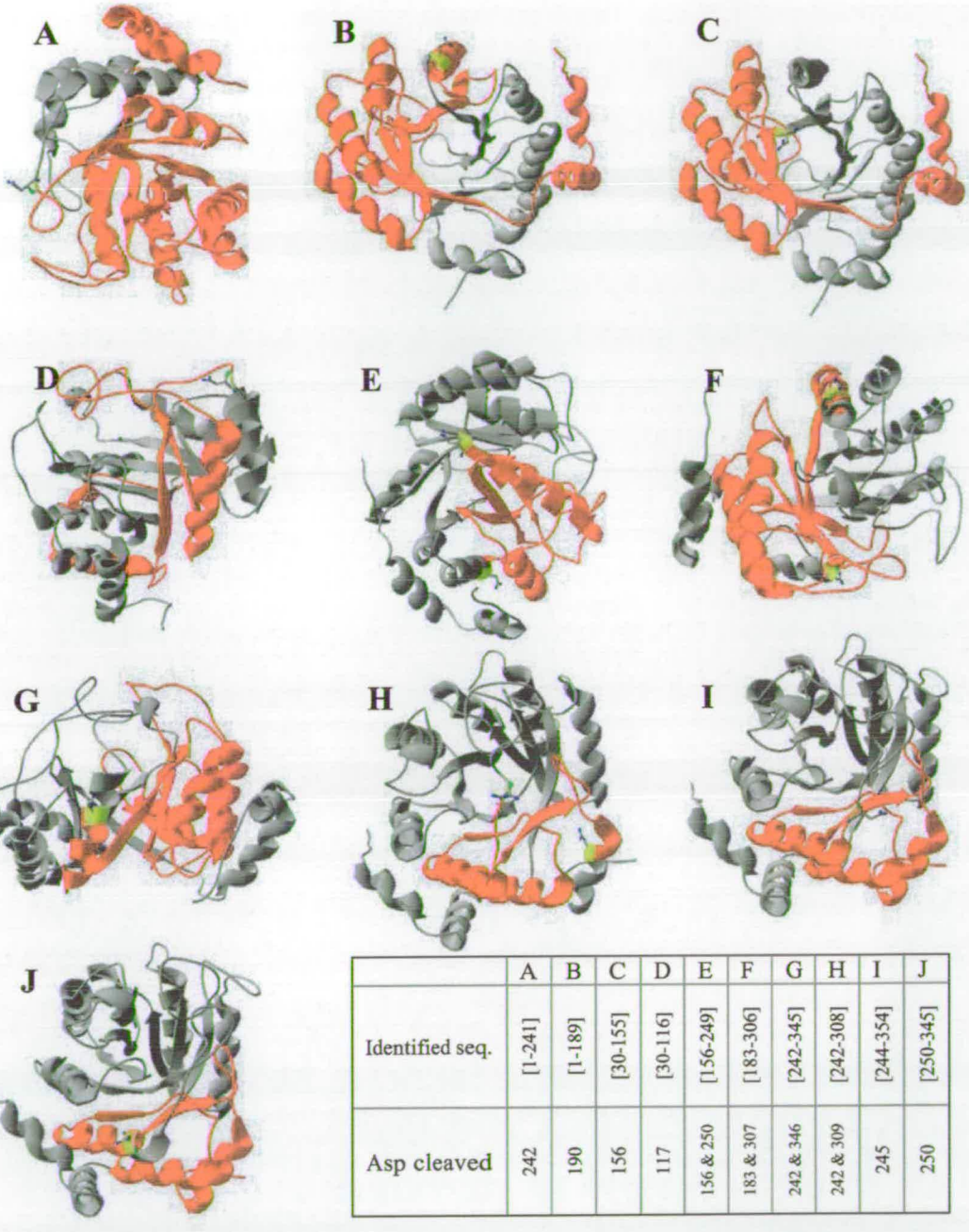
Digestion and Calibration of SDS-PAGE (fig 3.46)		HPLC purify and MALDI mass observation. (fig 3.48)		Dual Dig region mapping (fig 3.53)	
Band ID	Cal MW	Obs. Mass	Match region	Region Obs.	Consistent?
C	39.6kDa	38,445Da **38,662Da**	[1-343] **[1-345]**	 **[1-345]**	 ✓
E	29.1kDa	**26,975Da**	**[1-241]**	**[1-241]** [1-249]	✓ ✗
F	21.6kDa	**21,665Da**	**[1-189]**	**[1-189]** [209-343]	✓ ✗
G	18.3kDa	**17,593Da**	**[1-155]**	**[30-155]**	✓
H	13.1kDa	**13,454Da** 14,431Da **13,373Da**	**[1-116]** [117-249] **[183-306]**	**[30-116]** **[250-306]**	✓ ✓
I	11.4kDa	**10,292Da** **11,704Da** **12,486Da** **10,812Da**	**[156-249]** **[242-345]** **[244-354]** **[250-345]**	**[209-244]** **[250-343]** — ▪ — — ▪ —	✓ ✓ ✓ ✓
J	8.4kDa	7,450Da 9,894Da 8,305Da 9,400Da 8,480Da	[1-63] [30-116] [117-189] [156-241] [163-241]		✗ ✗ ✗ ✗

Figure 3.51 – Table showing summary of limited digestion analysis by Gel calibration, HPLC purification and MALDI and Dual digestion protocol. Data found to be consistent in all 3 analyses shown ticked in table. Selected sequences are shown mapped to crystal structure in Figure 3.55.

As a crystal structure of *E. coli* biotin synthase was published during the preparation of this manuscript a comparison of the identified fragments, and their associated cleavage sites was possible by mapping identified regions onto the crystal structure as shown in Figure 3.52 below.

Firstly, it is interesting to note that of the major species identified during the analysis all display the cleaved residues at surface exposed regions of the protein. Closer inspection of the mapped structures reveals an intriguing structural feature. Almost all structures in Figure 3.52, with the exception of E, tend to separate the structure into 2 sections best described in structure A. Intriguingly when this is compared to the crystal structures of SAM dependant enzymes (Figure 1.17) this roughly corresponds to the structurally conserved region of these 3 crystal structures. This may suggest that it is this larger N terminal section that is the reproduced in these enzymes. Firstly the main body of the protein, red in this case, consists of approximately residues 4 to 240 and a second section (grey in A) of approximately 240 to 354. An interesting structural feature is that the C-terminal section described seems to be sandwiched by the first 40 amino acids and the remaining protein folds. The first 40 amino acids seem to form a 'lid' of 2 helical sections linked to the main body by a flexible 8 residue loop. As most of the observed species originate from a cleavage releasing this C-term species from between this sandwich we hypothesise that this region may rotate around a backbone bond thus freeing the section. A theoretical example is shown in Figure 3.53.

Figure 3.53 shows the 2 identified sections of biotin synthase in Green (Main N-terminal 40 to 240 aa) and Red (C-terminal 240 to 354 aa), the 'lid' section in grey sitting on top of the C-terminal domain. A manipulation of the Q39–V40 peptide bond shows the shift of the whole 'lid' section from the surface of the proteins C-terminal end. The Asp rich region between D242-D247 is shown in yellow and cleavage at any of these residues post or pre rotation would yield the fragments [1-240] (See Figure 3.52 – A) and [241-354]. As the production of the [1-240] species was found consistently in all our analyses we proposed the production of a truncated gene encoding only this section of the protein and it complementary C-terminal partner to demonstrate the conformation of the distinct sections and flexibility of the first 40 amino acids.



	A	B	C	D	E	F	G	H	I	J
Identified seq.	[1-241]	[1-189]	[30-155]	[30-116]	[156-249]	[183-306]	[242-345]	[242-308]	[244-354]	[250-345]
Asp cleaved	242	190	156	117	156 & 250	183 & 307	242 & 346	242 & 309	245	250

Figure 3.52 – Identified limited digestion fragments of biotin synthase shown as part of the monomeric structure of published biotin synthase crystal structure. Observed fragment shown in red and cleaved Asp residue shown in yellow with side chain displayed.

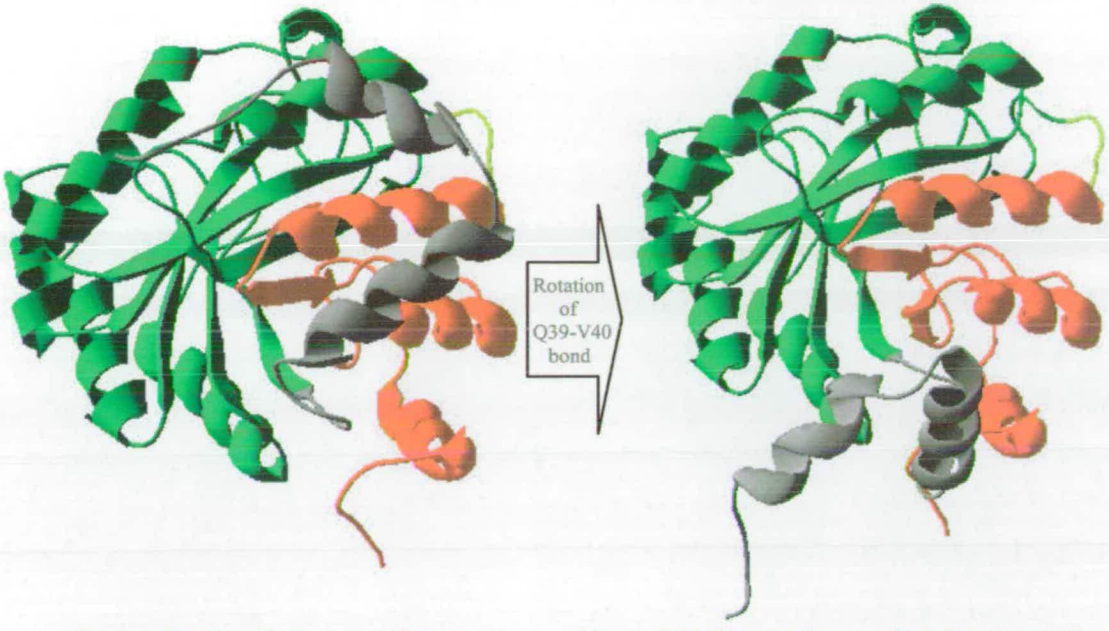


Figure 3.53 – X-Ray crystal structure of *E. coli* biotin synthase showing a rotation around Gln39–Val40 peptide bond that shifts the N-terminal 40 amino acids (grey) uncovering the C-terminal domain of the protein (red) for separation from remainder of protein (green) after cleavage at Asp rich loop shown in yellow.

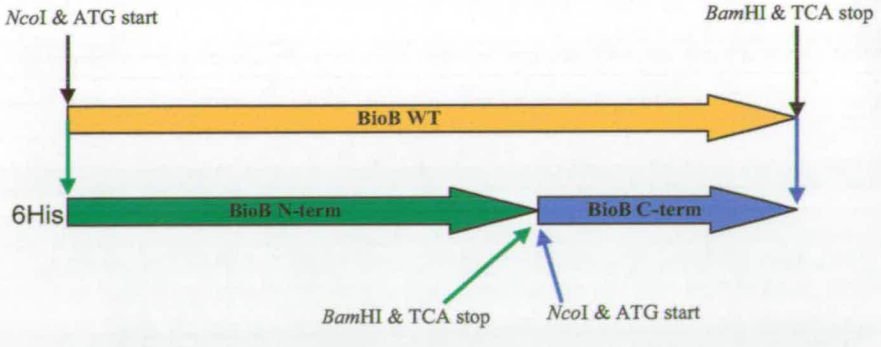


Figure 3.54 – Representation of BioB wild type gene and the truncated genes BioB_N-term and BioB_Cterm. Shows cloning sites and start/stop codons.

3.8.9 – Cloning and Expression of Derived Domains of Biotin Synthase

As in Section 3.1 pET16b/BioB was used as template DNA to produce the truncated biotin synthase genes. The gene was truncated removing the final 104 amino acids and also the corresponding fragment containing the final section of the native protein sequence as shown in Figure 3.54.

The truncated genes were amplified using standard PCR conditions and combination of the BioBfor, BioBrev, BSNterm_rev and BSCterm_for primers were used as required. The forward primer used in both cases contained a NcoI restriction site at N-term, containing the ATG methionine start codon and the reverse primers contain a BamHI site as well as encoding a TCA stop codon. This allowed both genes to be restriction digested with NcoI/BamHI combination and subsequently cloned into complementary digested plasmids. The BS_Nterm gene was cloned into pET6H to provide the N-term polyhis tag for purification and provides ampicillin resistance. The BS_C-term was cloned into the pET28a plasmid which provides the complementary kanamycin resistance, and no His tag, as represented in Figure 3.55.

Both plasmids were used to transform HMS174(DE3) *E. coli* cells and a single colony used to inoculate 2ml 2YT with appropriate antibiotic and protein production induced with addition of IPTG as standard conditions described in Section 2.16. After 3 hours of induction cells were harvested and analysed by SDS-PAGE. The calculated molecular weight of BS_N-term and BS-C-term are 27.5kDa and 12.3kDa respectively and band observed on SDS-PAGE was in agreement with the value of the BS_N-term as shown in Figure 3.56, however the band presumed to be BS_C-term was observed at 20kDa almost twice the expected value however sequencing of the gene confirmed that no elongation of the gene was seen and so the observed SDS-PAGE migration was attributed to the protein chain characteristics. When large scale protein induction was performed and cells disrupted with sonication the expressed proteins were found to be insoluble and present exclusively in the pellet after centrifugation. The protein was solubilised with Urea in increasing concentration from 2M to 10M and protein visualised in soluble fraction when a final concentration of 4M Urea was used. Subsequent attempts to refold proteins either on column or through serial dialysis yielded little or no observable soluble protein.

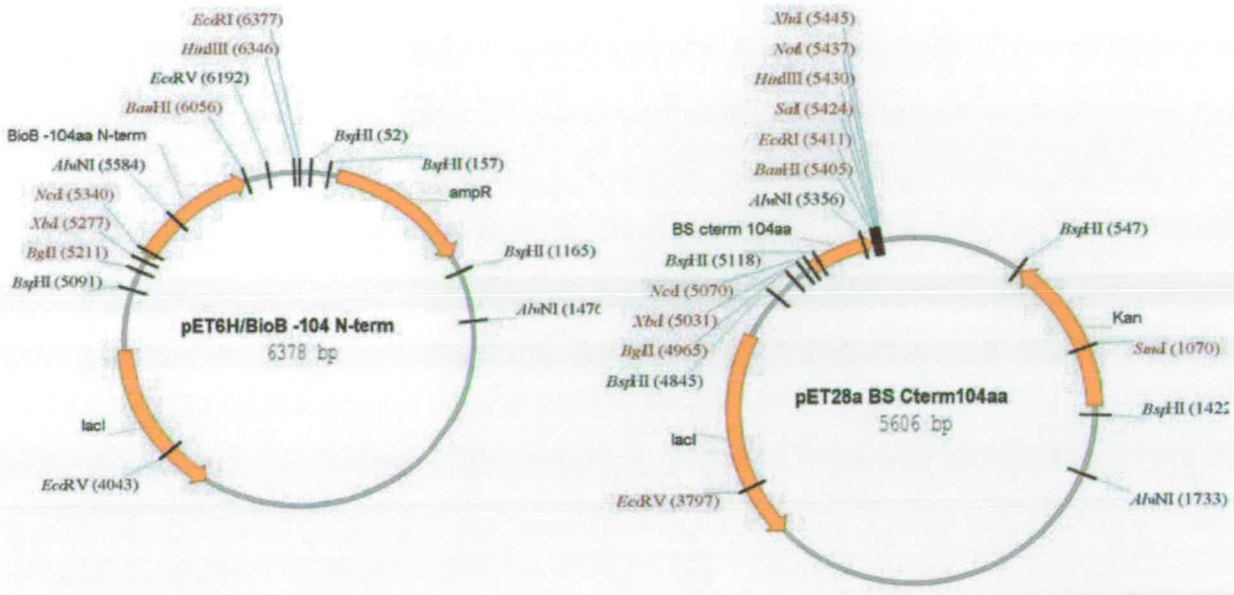


Figure 3.55 – Representation of the vectors used to express the two derived domains of biotin synthase. pET6H/BioB_Nterm encodes a Nterm 6 histidine tag to the protein whereas pET28a/BS_Cterm remains untagged. The vectors impart ampicillin and kanamycin resistance respectively to the transformed cells.

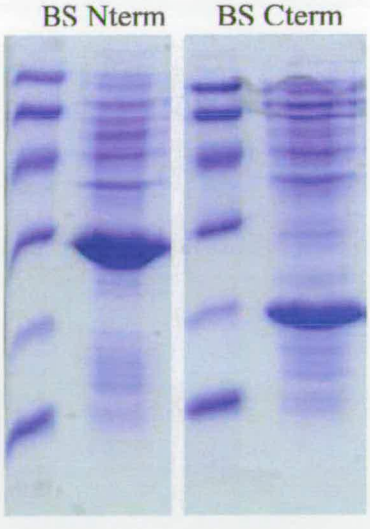


Figure 5.56 – SDS-PAGE of small scale induction of biotin synthase truncated proteins. Molecular weight marker used from Sigma, 98kDa – Phosphorylase B, 66kDa – BSA, 45kDa – Glutamic Dehydrogenase, 31kDa – Alcohol Dehydrogenase, 21kDa, - Myoglobin, 14kDa – Lysosyme.

During the design of the experiment the choice of plasmids was crucial to allow transformation of cells with both plasmids and selection for both using their complementary resistance genes, ampicillin and kanamycin. A transformation of HMS174(DE3) with both plasmids was achieved using 2µl of each plasmid DNA solution and overnight growth on LB(amp/kan) plate however transformation efficiency was very low with typically only 2-3 colonies found per transformation reaction. These colonies were subjected to small scale induction as before and whole cells analysed by SDS-PAGE, and the gel is shown in Figure 3.57 showing induced and control growths. When large scale induction was performed using conditions previously described for 6His-BS approximately 10% of both expressed proteins were found to be soluble and CFE displayed the characteristic red/brown colour as observed during previous preparations of BS and mutants.

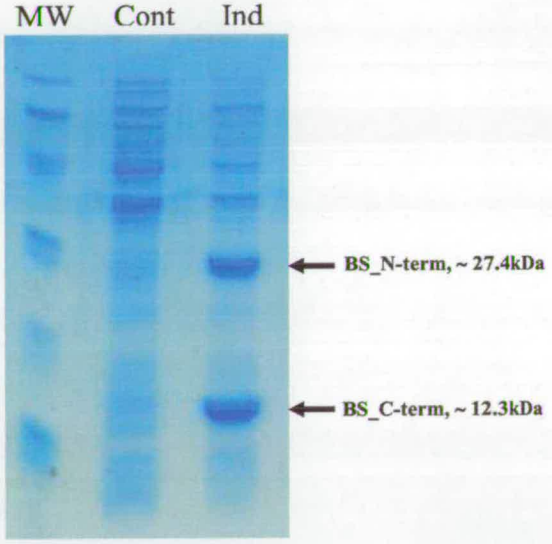


Figure 5.57 – SDS-PAGE of small scale dual induction of biotin synthase truncated proteins expressed in same cell. Molecular weight marker used from Sigma, 98kDa – Phosphorylase B, 66kDa – BSA, 45kDa – Glutamic Dehydrogenase, 31kDa – Alcohol Dehydrogenase, 21kDa, - Myoglobin, 14kDa – Lysosyme.

3.8.10 – Purification and Characterisation of Biotin Synthase Domain Complex

The CFE from cells expressing both 6His-BS_N-term and BS_C-term was prepared using BufferA for loading onto a Ni-HiTrap column from Amersham-Pharmacia as described for the intact WT biotin synthase and mutants. When CFE was loaded through the column, a red/brown band formed and remained during the washing procedure. The band was eluted at approximately 200 mM imidazole concentration and yielded three 5ml fractions red in colour. The elution fractions were analysed by SDS-PAGE and fractions 13, 14 and 15 that were red in colour showed presence of two co-eluting bands with observed molecular weight of approximately 28 kDa and 15 kDa which corresponded to the two expressed protein fragments as shown in Figure 3.58. This purification was performed to purify the 6His tagged N-term fragment of biotin synthase however the dual expression of the fragments has achieved formation of a complex which has been maintained during the column purification and the non tagged C-term fragment is strongly associated with the tagged fragment and co-purification was achieved. Upon closer examination of the crystal structure it can clearly been proposed that an N-term loop/helix folds down over the C-term fragment as seen in Figure 3.53. As seen in the SDS-PAGE analysis the complex is destroyed by the conditions used for the gel thus yielding the dissociated fragments as expected from the linear sequences.

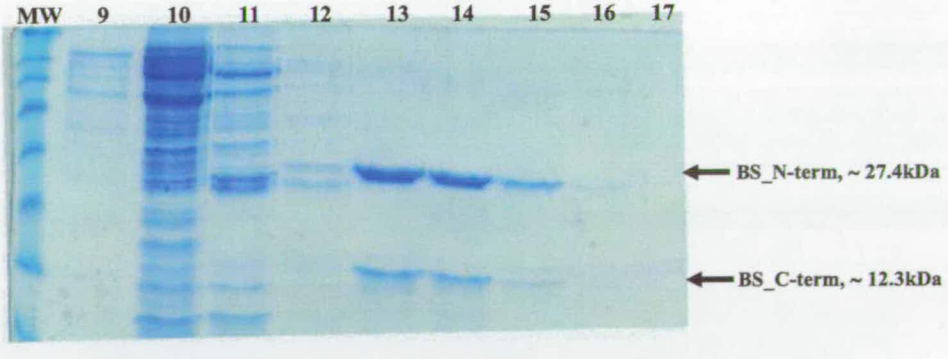


Figure 5.58 – SDS-PAGE analysis of purification fractions of biotin synthase truncated proteins expressed in same cell showing co-elution from Ni-HiTrap column. Molecular weight marker used from Sigma, 98kDa – Phosphorylase B, 66kDa – BSA, 45kDa – Glutamic Dehydrogenase, 31kDa – Alcohol Dehydrogenase, 21kDa, - Myoglobin, 14kDa – Lysosyme.

The purified complex seemed to exhibit similar characteristics to the intact protein and so simple characterisation of the complex was performed to compare this with 6His-BS intact proteins. UV-Vis analysis as shown in Figure 3.59 demonstrates that the complex displays maxima similar as 6His-BS but shows larger ratios between both 330:280 nm and 425:280 nm compared to 6His-BS. This suggests that this complex is a viable [Fe-S] cluster binding species and the UV-Vis evidence suggests that this species could be similar to that found in the as purified biotin synthase protein, namely one or two [2Fe-2S] per monomer.

Conformation of the components, thought to be the N/C-term complex, of the Fe binding truncated mutant was sought with utilisation of our standard LC-MS conditions as used previously for simple characterisation of the WT and mutant proteins. Due to the formation of the strong complex of the two fragments of biotin synthase it was unclear as to its behaviour during our LC-MS conditions. However as discussed in Section 3.4 we have evidence to suggest that 6HisWT biotin synthase and many mutants display an ion-envelope in which we observe a bias toward lower charge states and thus may indicate some existence of some tertiary structure maintained during LC-MS conditions. This being the case our observation of the complex as a single entity during LC-MS conditions, although surprising, does tend to validate our earlier observations are consistent with the assumptions made. The spectrum obtained during the LC-MS analysis was transformed and the mass spectra are shown in Figure 3.60. The obtained mass is consistent with the calculated mass of complex as shown in the table. The ion-envelope shows reasonably well defined low charge species (high m/z) but becomes increasingly poor in resolution toward the higher charge states (low m/z). The combination of these factors could contribute to the 500 Da at 50% peak width and complex multi species transformation peak in **B**. Important factors also affecting the quality of the obtained spectra could be the low absolute concentration of the observed complex due to several non specific complexes and loss of a portion of the complex during analysis and so some on column precipitation. Formation of adducts either non-specific (buffer ions etc.) of more pertinent modifications of Biotin synthase formed *in vivo* will also complicate spectra but due to the low resolution further investigation of these is impossible from these obtained spectra.

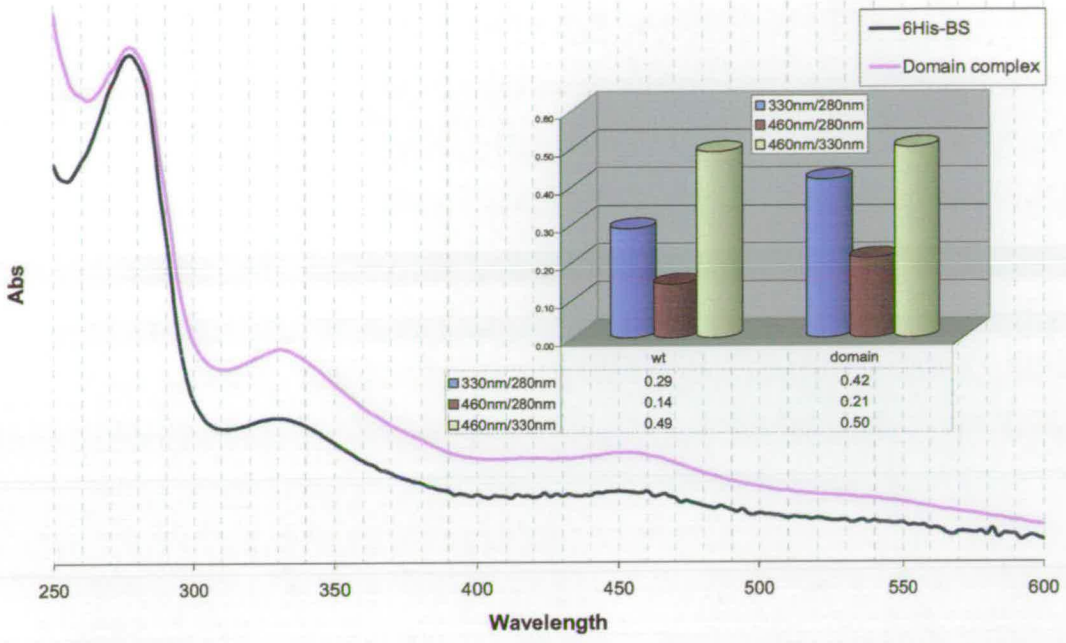
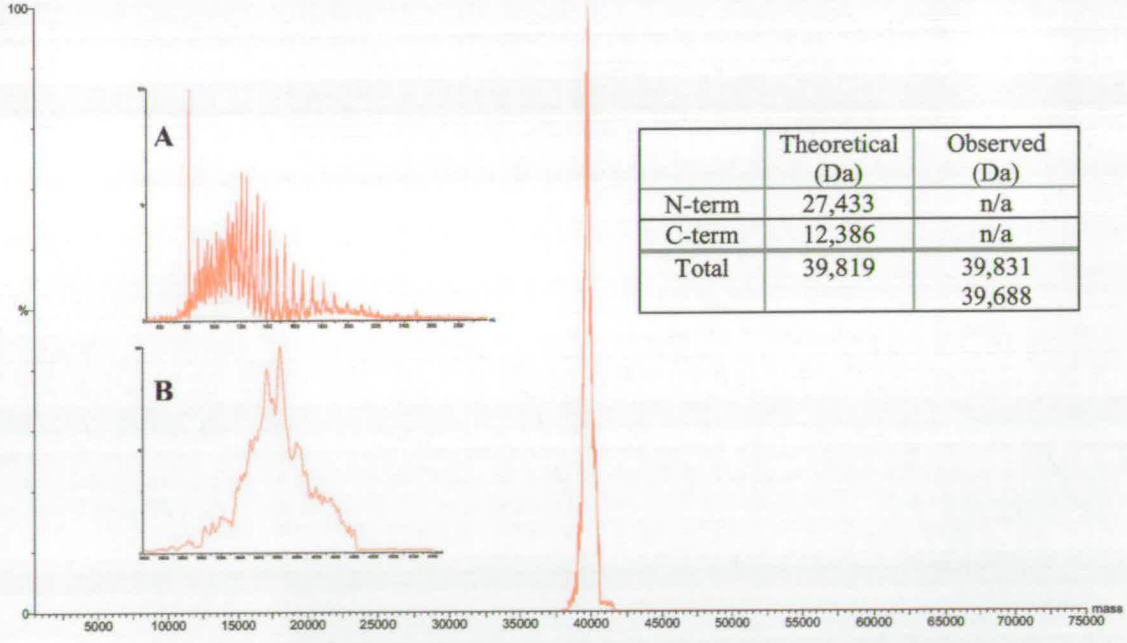


Figure 3.59 – UV-Vis scan from 250nm to 600nm of Domain complex and 6HisWT biotin synthase both at 1mg/ml concentration. Bar chart inset highlights ratios between 280:330, 280:425 and 330:425 wavelengths.



	Theoretical (Da)	Observed (Da)
N-term	27,433	n/a
C-term	12,386	n/a
Total	39,819	39,831 39,688

Figure 3.60 – Transformed Mass Spectrum observed after LC-MS analysis of domain complex showing maintenance of complex throughout LC-MS conditions and confirming molecular weight of complex compared to calculated protein mass, shown in table. Inset A = Ion envelope used for transformation. Inset B = Zoomed transformed peak showing main 39,831Da and 39,688Da peaks and other indefinable mass species.

References

References

1. Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T., and Throck Watson, J. (1987) *Anal. Chem.* **59**, 2621
2. Alexeev, D. (1999)
3. Frazzon, J., and Dean, D. R. (2003) *Curr Opin Chem Biol* **7**, 1-8
4. Wildiers, E. (1901) *Lacellute* **18**, 313-
5. Marsh, E. N. G., Patwardhan, A., and Huhta, M. S. (2004) *Bioorganic Chemistry* **32**, 326-340
6. Zheng, L., White, R. H., Cash, V., and Dean, D. (1994) *Biochem.* **33**, 4714-4720
7. Ugulava, N. B., Sacanell, C. J., and Jarrett, J. T. (2001) *Biochemistry* **40**, 8352-8358
8. Duin, E. C., Lafferty, M. E., Crouse, B. R., Allen, R. M., Sanyal, I., Flint, D. H., and Johnson, M. K. (1997) *Biochem.* **36**, 11811-11820
9. Fontecave, M., Ollagnier-de-Choudens, S., and Mulliez, E. (2003) *Chemical Reviews* **103**, 2149-2166
10. Wilson, K. P., Shewchuk, L. M., Brennan, R. G., Otsuka, A. J., and Matthews, B. W. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9257-9261
11. Kaltashov, I. A., and Eyles, S. J. (2002) *Mass Spectrometry Reviews* **21**, 37-71
12. Alexeev, D., Alexeeva, M., Baxter, R. L., Campopiano, D. J., Webster, S. P., and Sawyer, L. (1998) *J. Mol. Biol.* **284**, 401-419
13. Cupp-Vickery, J. R., Urbina, H., and Vickery, L. E. (2003) *Journal of Molecular Biology* **330**, 1049-1059
14. Hanzelmann, P., and Schindelin, H. (2004) *PNAS* **101**, 12870-12875
15. Alexeev, D., Bury, S. M., Boys, C. W. G., Turner, M. A., Sawyer, L., Ramsey, A. J., Baxter, H. C., and Baxter, R. L. (1994) *Journal of Molecular Biology* **235**, 774-776
16. Berkovitch, F., Nicolet, Y., Wan, J. T., Jarrett, J. T., and Drennan, C. L. (2004) *Science* **303**, 76-79
17. Williams, R., J, Lyman, C. M., Goodyear, G. M., Truesdail, H. J. H., and Holaday, D. (1943) *Journal of Amercian Chemical Society* **55**, 2912-
18. Kögl, F., and Tonnis, B. (1936) *Z. Physiol Chemistry* **242**, 43-73
19. Du Vigneaud, V., Melville, D. B., Moyer, A. W., and Hofmann, K. (1942) *Journal of Biological Chemistry* **146**, 475
20. Goldberg, M. W., and Sternbach, H. (1946), United States Patent
21. McGarrity, J., TTenud, L., and Meul, T. (1988), European Patent Office

22. Lynen, F., Knappe, J., Lorch, E., Jutting, G., and Rungelman, E. (1959) *Agnew Chemistry* **71**, 481-486
23. Tipton, P. A., and cleland, W. W. (1988) *Biochem.* **27**, 4317-4325
24. Ogita, T., and Knowles, J. R. (1988) *Biochem.* **27**, 8028-8033
25. Eisenberg, M. A. (1973) *Advances in Enzymology Related Areas Molecular Biology* **38**, 317-372
26. Moss, J., and Lane, M. D. (1971) *Advances in Enzymology* **35**, 321-442
27. Kasow, D. P., and Lane, M. D. (1962) *Biochemistry and Biophysical Research Communications* **7**, 439-443
28. Hillier, Y., Gershoni, J. M., Bayer, E. A., and Wilchek, M. (1987) *Biochemistry Journal* **248**, 167-171
29. Wilchek, M., and Bayer, E. A. (1988) *Analytical Biochemistry* **171**, 1-32
30. Lindqvist, Y., and Schneider, G. (1996) *Curr. Opinion in Struct. Biol.* **6**, 798-803
31. Whitehead, C. C. (1988) *Biotin in animal nutrition*, Hoffmann La-Roche, Basel
32. Swick, H. M., and Klein, C. L. (1985) *Ann New York Academy of Science* **447**, 430
33. Taylor, A. L. (1970) *Bacteriological Reviews* **34**, 155-175
34. del Campillo Campbell, A., Kayajanian, G., Campbell, A., and Adhya, S. J. (1967) *J. Bact.* **94**, 2065-2066
35. Rolfe, B., and Eisenberg, M. A. (1968) *J. Bact.* **96**, 515-524
36. Cleary, P. P., and Campbell, A. (1972) *J. Bact.* **112**, 830-839
37. Rolfe, B. (1970) *Virology* **42**, 643-661
38. Guha, A., Saturen, Y., and Szybalski, W. (1971) *Journal of Molecular Biology* **56**, 53-62
39. Eisenberg, M. A. (1975) *Metabolic Pathways*, 27, Academic Press, New York
40. Barker, D. F., Kuhn, J., and Campbell, A. M. (1981) *Gene* **13**, 89-102
41. Pai, C. H., and Lichstein, H. C. (1965) *Biochimie and Biophysica Acta* **100**, 28-35
42. Eisenberg, M. A., and Star, C. (1968) *J. Bact.* **96**, 1291-1297
43. Eisenberg, M. A., Prakesh, O., and Hsuing, S. (1982) *J. Biol. Chem.* **257**, 15167-15173
44. Howard, P. K., Shaw, J., ,, and Otsuka, A. (1985) *Gene* **35**, 321-331
45. Weaver, L. H., Kwon, K., Becket, D., and Matthews, B. W. (2001) *PNAS* **98**, 6045-6050
46. Ploux, O., Soularue, P., Marquet, A., Gloeckler, R., and Lemoine, Y. (1992) *Biochemical Journal* **287**, 685-690

47. Sanyal, I., Lee, S. L., and Flint, D. H. (1994) *Journal of the American Chemical Society* **116**, 2637-2638
48. Ifuku, O., Miyaoka, H., Koga, N., Kishimoto, J., Haze, S., Wachi, Y., and Kajiwara, M. (1994) *European Journal of Biochemistry* **220**, 585-591
49. Lemoine, Y., Wach, A., and Jeltsch, J. M. (1996) *Molecular Microbiology* **19**, 645-647
50. Tomczyk, N. H., Nettleship, J. E., Baxter, R. L., Crichton, H. J., Webster, S. P., and Campopiano, D. J. (2002) *Febs Letters* **513**, 299-304
51. Sanishvili, R., Yakunin, A. F., Laskowski, R. A., Skarina, T., Evdokimova, E., Doherty-Kirby, A., Lajoie, G. A., Thornton, J. M., Arrowsmith, C. H., Savchenko, A., Joachimiak, A., and Edwards, A. M. (2003) *Journal of Biological Chemistry* **278**, 26039-26045
52. Ploux, O., and Marquet, A. (1992) *Biochemical Journal* **283**, 327-331
53. Eliot, A. C., and Kirsch, J. F. (2004) *Annual Review of Biochemistry* **73**, 383-415
54. Webster, S. P., Alexeev, D., Campodiano, D. J., Watt, R. M., Alexeeva, M., Sawyer, L., and Baxter, R. L. (2000) *Biochemistry* **39**, 516-528
55. Kack, H., Sandmark, J., Gibson, K., Schneider, G., and Lindqvist, Y. (1999) *J. Mol. Biol.* **291**, 857-876
56. Sandmark, J., Mann, S., Marquet, A., and Schneider, G. (2002) *Journal of Biological Chemistry* **277**, 43352-43358
57. Krell, K., and Eisenberg, M. A. (1970) *J. Biol. Chem.* **245**, 6558-6566
58. Sakurai, N., Imai, Y., Masuda, M., Komatsubara, S., and Tosa, T. (1993) *Appl. Environ. Microbiol.* **59**, 2857-2863
59. Gloeckler, R., Ohsawa, I., Speck, D., Ledoux, C., Bernard, S., Zinsius, M., Villeval, D., Kisou, T., Kamogawa, K., and Lemoine, Y. (1990) *Gene* **87**, 63-70
60. Huang, W. J., Lindqvist, Y., Schneider, G., Gibson, K. J., Flint, D., and Lorimer, G. (1994) *Structure* **2**, 407-414
61. Baxter, R. L., Ramsey, A. J., McIver, L. A., and Baxter, H. C. (1994) *J. Chem. Soc. Chem. Commun.*, 559-560
62. Baxter, R. L., and Baxter, H. C. (1994) *J. Chem. Soc. Chem. Commun.*, 759-760
63. Gibson, K. J., Lorimer, G. H., Rendina, A. R., Taylor, W. S., Cohen, G., Gatenby, A. A., Payne, W. G., Roe, D. C., Lockett, B. A., Nudelman, A., Marcovici, D., Nachum, A., Wexler, B. A., Marsilii, E. L., Turner, I. M., Howe, L. D., Kalbach, C. E., and Chi, H. J. (1995) *Biochemistry* **34**, 10976-10984

64. Gibson, K. J., Lorimer, G. H., Rendina, A. R., Taylor, W. S., Cohen, G., Gatenby, A. A., Payne, W. G., Roe, D. C., Lockett, B. A., and Nudelman, A. (1995) *Biochemistry* **34**, 10976-10984
65. Florentin, D., Bui, B. T. S., Marquet, A., Ohshiro, T., and Izumi, Y. (1994) *Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences* **317**, 485-488
66. Baldet, P., Alban, C., and Douce, R. (1997) *Febs Letters* **419**, 206-210
67. Flint, D. H., and Allen, R. M. (1997) *Meth. Enzym.* **279**, 349-356
68. Sanyal, I., Cohen, G., and Flint, D. H. (1994) *Biochemistry* **33**, 3625-3631
69. Serebriiskii, I. G., Vassin, V. M., and Tsygankov, Y. D. (1996) *Gene* **175**, 15-22
70. Bower, S., Perkins, J. B., Yocum, R. R., Howitt, C. L., Rahaim, P., and Pero, J. (1996) *Journal of Bacteriology* **178**, 4122-4130
71. Wu, C. H., Chen, H. Y., and Shiuan, D. (1996) *Gene* **174**, 251-258
72. Otsuka, A. J., Buonovistiani, M. R., Howard, P. K., Flamm, J., Johnson, C., Yamamoto, R., Uchida, K., Cook, C., Ruppert, J., and Matsuzaki, J. (1988) *J. Biol. Chem.* **263**, 19577-19585
73. Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F., and Miller, N. E. (2001) *Nucleic Acids Research* **29**, 1097-1106
74. Ollangier, S., Meier, C., Mulliez, E., Gaillard, J., Schuenemann, V., Trautwein, A., Mattioloi, T., Lutz, M., and Fontecave, M. (1999) *J.Am.Chem. Soc.* **121**, 6344-6350
75. Leuthner, B., Leutwin, C., Schulz, H., Horth, P., Haehnel, W., Sclitz, E., Schagger, H., and Heider, J. (1998) *Mol. Microbiol.* **28**, 615-628
76. Ollagnier-de Choudens, S., and Fontecave, M. (1999) *Febs Letters* **453**, 25-28
77. Frey, P. A., and Reed, G. H. (1993) *Adv. Enzymolo. Rel. Areas Mol. Biol.* **66**
78. Joerger, R. D., and Bishop, P. E. (1998) *J. Bacteriol.* **170**, 1475-1487
79. Shaw, D. J., and Guest, J. R. (1982) *Nucleic Acids Res.* **10**, 6119-6130
80. Liu, S. T., Lee, L., TAi, C. Y., Hung, C. H., Chang, Y. S., Wolfram, J. H., Rogers, R., and Goldstein, A. H. (1992) *J. Bacteriol.* **174**, 5814-5819
81. Broderick, J. B., Duderstadt, R. E., Fernandez, D. C., Wojtuszewski, K., Henshaw, T. F., and Johnson, M. K. (1997) *Journal of American Chemical Society* **119**, 7396-7397
82. Fajardo-Cavazos, P., Salazar, C., and Nicholson, W. L. (1993) *J. Bacteriol.* **175**, 1735-1744
83. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H. M., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qui, D., Spadafora, R.,

Vicare, R., Wang, Y., Weizerbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrovski, S., Church, G. M., Daniels, C. J., Mao, J. I., Rice, P., Nolling, J., and Reeve, J. N. (1997) *J. Bacteriol.* **179**

84. Glaser, P., Danchin, A., Kunst, F., Zuber, P., and Nakano, M. (1995) *J. Bacteriol.* **177**, 1112-1115
85. Martinez-Gomez, N. C., Robers, M., and Downs, D. M. (2004) *J Biol Chem* **279**, 40505-40510
86. Hewitson, K. S., Baldwin, J. E., Shaw, N. M., and Roach, P. L. (2000) *FEBS letts* **466**, 372-376
87. McIver, L., Baxter, R. L., and Campopiano, D. J. (2000) *Journal of Biological Chemistry* **275**, 13888-13894
88. Tse Sum Bui, B., Florentin, D., Marquet, A., Benda, R., and Trautwein, A. X. (1999) *FEBS letts* **459**, 411-414
89. Ollagnier-de Choudens, S., Sanakis, Y., Hewitson, K. S., Roach, P., Baldwin, J. E., Munck, E., and Fontecave, M. (2000) *Biochemistry* **39**, 4165-4173
90. Ugulava, N. B., Gibney, B. R., and Jarrett, J. T. (2000) *Biochemistry* **39**, 5206-5214
91. Ollagnier-de-Choudens, S., Mulliez, E., and Fontecave, M. (2002) *Febs Letters* **532**, 465-468
92. Ollagnier-de-Choudens, S., Mulliez, E., Hewitson, K. S., and Fontecave, M. (2002) *Biochemistry* **41**, 9145-9152
93. Ollagnier-de Choudens, S., Sanakis, Y., Hewitson, K. S., Roach, P., Munck, E., and Fontecave, M. (2002) *Journal of Biological Chemistry* **16**, 13449-13454
94. Cosper, M. M., Jameson, G. N. L., Eidsness, M. K., Huynh, B. H., and Johnson, M. K. (2002) *FEBS Letters* **529**, 332-336
95. Sanyal, I., Gibson, K. J., and Flint, D. H. (1996) *Arch. Biochem. Biophys.* **326**, 48-56
96. Birch, O. M., Fuhrmann, M., and Shaw, N. M. (1995) *Journal of Biological Chemistry* **270**, 19158-19165
97. Ohshiro, T., Yamamoto, M., Izumi, Y., Bui, B. T. S., Florentin, D., and Marquet, A. (1994) *Bioscience Biotechnology and Biochemistry* **58**, 1738-1741
98. Ifuku, O., Kishimoto, J., Haze, S., Yanagi, M., and Fukushima, S. (1992) *Bioscience Biotechnology and Biochemistry* **56**, 1780-1785
99. Ohshiro, T., Kishimoto, J., Arase, M., and Izumi, Y. (1998) *Journal of bioengineering* **217**, 1231

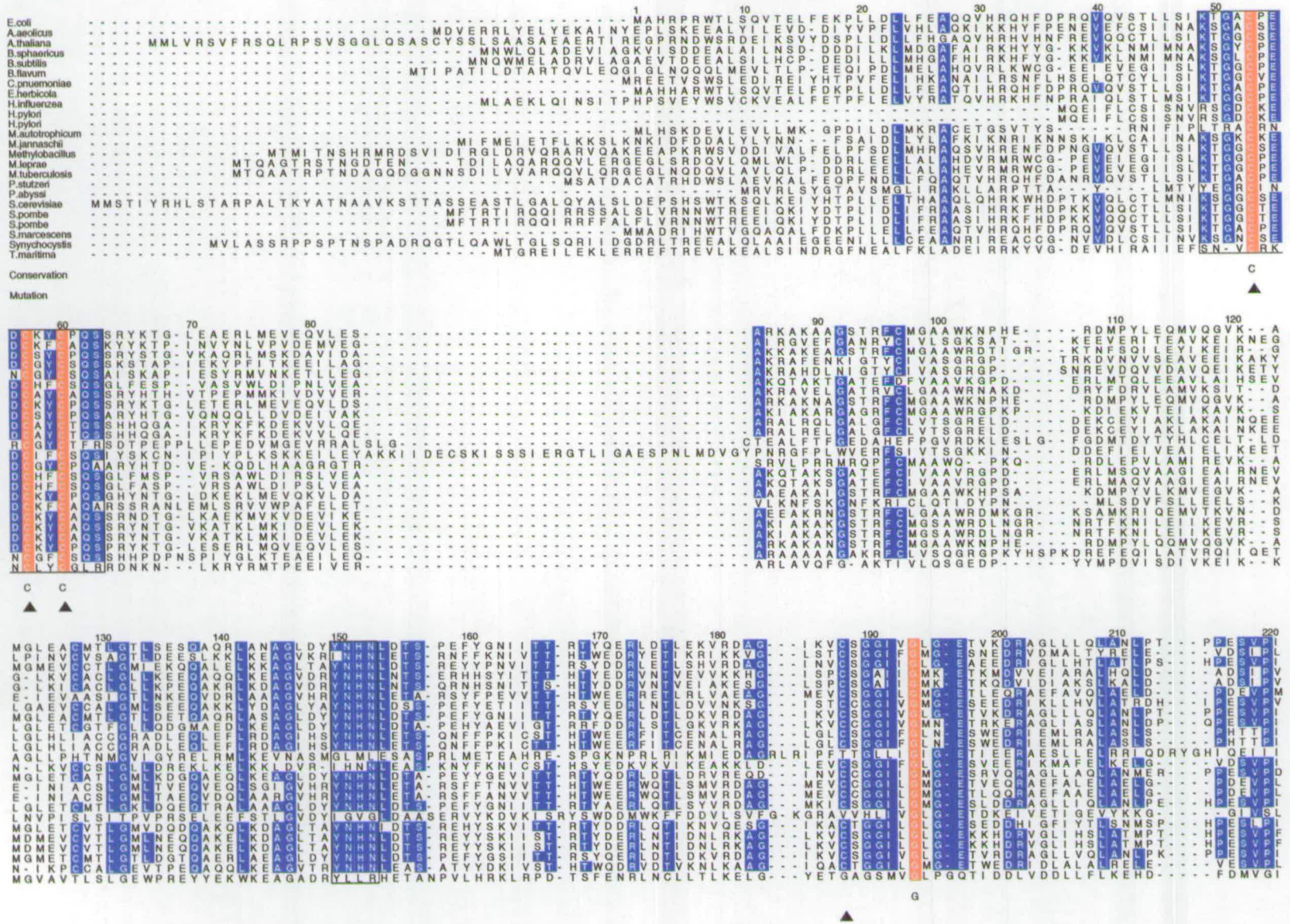
100. Guianvarc'h, D., Florentin, D., Tse Sum Bui, B., Nunzi, F., and Marquet, A. (1997) *Biochem. Biophys. Res.Comm.* **236**, 402-406
101. Escalettes, F., Florentin, d., Tse Sum Bui, B., Lesage, D., and Marquet, A. (1999) *J. Am Chem. Soc.* **121**, 3571-3578
102. Shaw, N. M., Birch, O. M., Tinschert, A., Venetz, V., Dietrich, R., and Savoy, L. A. (1998) *Biochemical Journal* **330**, 1079-1085
103. Jarrett, J. T. (2005) *Chemistry and Biology* **12**, 409-415
104. Tse Sum Bui, B., Escalettes, F., Chottard, G., Florentin, D., and Marquet, A. (2000) *Eur. J., Biochem.* **267**, 2688-2694
105. Kiyasu, T., Asakura, A., Nagahashi, Y., and Hoshino, T. (2000) *J. Bact.* **182**, 2879-2885
106. Tse Sum Bui, B., Florentin, D., Fournier, F., Ploux, O., Mejean, A., and Marquet, A. (1998) *FEBS Letts* **440**, 226-230
107. Tse Sum Bui, B., Lotierzo, M., Escalettes, F., Florentin, D., and Marquet, A. (2004) *Biochemistry* **43**, 16432-16441
108. Jarrett, J. T. (2003) *Current Opinion in Chemical Biology* **7**, 174-182
109. Hewitson, K. S., Ollagnier-de Choudens, S., Sanakis, Y., Shaw, N. M., Baldwin, J. E., Munck, E., Roach, P. L., and Fontecave, M. (2002) *Journal of Biological Inorganic Chemistry* **7**, 83-93
110. Murzin, A. G. (1995) *J. Mol. Biol.* **247**, 536
111. Ugulava, N. B., Frederick, K. K., and Jarrett, J. T. (2003) *Biochem.*, 2708-2719
112. Leider, K. W., Booker, S. J., Ruzicka, F. J., Beinert, H., Reed, G. H., and Frey, P. A. (1998) *Biochemistry* **37**, 2578-2585
113. Becker, A., Fritzwolf, K., Kabsch, W., knappe, J., Schultz, S., and Wagner, A. F. (1999) *Nat. Struct. Biol.* **6**, 969-975
114. Leppanen, V. M., Merckel, M. C., Ollis, D. L., Wong, K. K., Kazrich, J. W., and Goldman, A. (1999) *Structure with Folding and Design* **7**, 733-744
115. Volker Wagner, A. F., Frey, M., Neugebauer, F. A., and Schafer, W. (1992) *Proceedings of the National Academy of Science USA* **89**, 996-1000
116. Conradt, H., Hohmann-Berger, M., Hohmann, H., Blaschowski, H. P., and Knappe, J. (1984) *Archives of Biochemistry and Biophysics* **28**, 133-142
117. Wong, K. K., Murray, B. W., Lewisch, S. A., Baxter, M. K., Ridky, T. W., Ulissi-DeMario, L., and Kozarich, J. W. (1993) *Biochemistry* **32**, 14102-14110
118. Kulzer, R., Pils, T., Kappl, R., Huttermann, J., and Knappe, J. (1998) *Journal of Biological Chemistry* **273**, 4897-4903

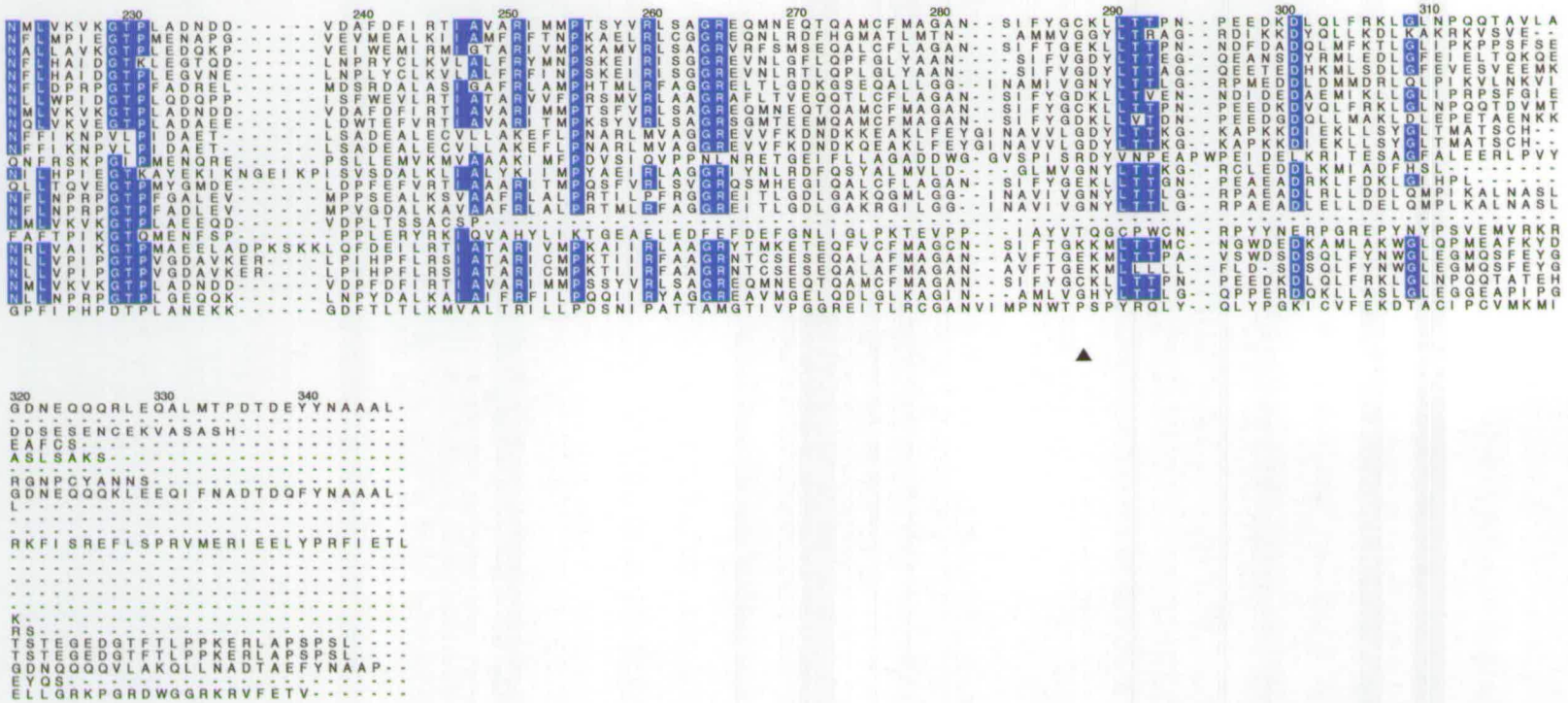
119. Sun, X., Ollagnier, S., Schmidt, P. P., Atta, M., Mulliez, E., Lepape, L., Eliasson, R., Graslund, A., Fontecave, M., Reichard, P., and Sjoberg, B. (1996) *Journal of Biological Chemistry* **271**, 6827-6831
120. Mulliez, E., Fontecave, M., Gaillard, J., and Reichard, P. (1993) *J. Biol. Chem.* **268**, 2296-2299
121. Busby, R. W., Schelvis, J. P. M., Yu, D. S., Babcock, G. T., and Marletta, M. A. (1999) *Journal of the American Chemical Society* **121**, 4706-4707
122. Miller, J. R., Busby, R. W., Jordan, S. W., Cheek, J., Henshaw, T. F., Ashley, G. W., Broderick, J. B., Cronan, J. E., and Marletta, M. A. (2000) *Biochemistry* **39**, 15166-15178
123. Morris, T. W., Reed, K. E., and Cronan, J. E. (1995) *J. Biol. Chem* **177**, 1-10
124. Jordan, S. W., and Cronan, J. E. (2003) *J. Bacteriol* **185**, 1582-1589
125. Zhao, X., Miller, J. R., Jiang, Y., Marletta, M. A., and Cronan, J. E. (2003) *Chem. Biol.* **10**, 1293-1302
126. Bryant, P., Kriek, M., Wood, R. J., and Roach, P. (2005) *Anal. Biochem.* **351**, 44-49
127. Cicchillo, R. M., and Booker, S. (2005) *J Am Chem Soc* **127**, 2860-2861
128. Hanzelmann, P., Schwartz, G., and Mendel, R. R. (2002) *Journal of Biological Chemistry* **277**, 18303-18312
129. Mortenson, L. E., Valentine, R. C., and Carnahan, J. E. (1962) *Biochem, Biophys. Res. Comm.* **7**, 448-452
130. Johnson, M. K. (1998) *Curr. Opin. Chem. Biol.* **2**, 173-181
131. Beinert, H., Hol, H. R., and Munck, E. (1997) *Science* **277**, 653-659
132. Wachtershauser, G. (1992) *Prog Biophys Mol Biol* **58**, 85-201
133. Einsle, O., Tezcan, F. A., Andrade, S. L. A., Schmid, B., Yoshida, M., Howard, J. B., and Rees, D. C. (2002) *Science* **297**, 1696-1700
134. Hagen, K. S., Reynolds, J. G., and Holm, R. H. (1981) *J Am Chem Soc* **103**, 4054-4063
135. Nakamura, M., Saeki, K., and Takahashi, Y. (1999) *J Biochem* **126**, 10-18
136. Schwartz, C. J., Djaman, O., Imlay, J. A., and Kiley, P. J. (2000) *Proceedings of the National Academy of Science USA* **97**, 9009-9014
137. Evans, D. J., Jones, R., Woodley, P. R., Wilborn, J. R., and Robson, R. L. (1991) *J Bacteriol* **173**, 5457-5469
138. Schwartz, C. J., Giel, J. L., Patschowski, T., Luther, C., Ruzicka, F. J., Beinert, H., and Kiley, P. J. (2001) *Proceedings of the National Academy of Science USA* **98**, 14895-14900
139. Lauhon, C. T., and Kambampti, R. (2000) *J Biol Chem* **275**, 20096-22013
140. Zheng, L., White, R. H., Cash, V. L., and Dean, D. R. (1994) *Biochemistry* **33**, 4714-4720

141. Ollagnier-de-Choudens, S., Mattiol, T., Takahashi, T., and Fontecave, M. (2001) *J. Bio. Chem.* **276**, 22604-22607
142. Agar, J. N., Zheng, L., Cash, V. L., Dean, D. R., and Johnson, M. K. (2000) *J Am Chem Soc* **122**, 7856-7863
143. Hoff, K. G., Silberg, J. J., and Vickery, L. E. (2000) *Proceedings of the National Academy of Science USA* **97**, 7790-7795
144. Jacobson, M. R., Cash, V. L., Weiss, M. C., Laird, N. F., Newton, W. E., and Dean, D. R. (1989) *Molecular & General Genetics* **219**, 49-57
145. Zheng, L. M., White, R. H., Cash, V. L., Jack, R. F., and Dean, D.R.,. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 2754-2758
146. Mihara, H., and Esaki, N. (2002) *Applied Microbiology and Biotechnology* **60**, 12-23
147. Mihara, H., Maeda, M., Fujii, T., Kurihara, T., Hata, Y., and Esaki, N. (1999) *J Biol Chem* **274**, 14768-14772
148. Mihara, H., Kurihara, T., Yoshimura, T., and Esaki, N. (2000) *J Biochem (Tokyo)* **127**, 559-567
149. Urbina, H. D., Silberg, J. J., Hoff, K. G., and Vickery, L. E. (2001) *J Biol Chem* **276**, 44521-44526
150. Aebersold, R., and Mann, M. (2003) *Nature* **422**, 198-207
151. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) *Science* **246**, 64-71
152. Karas, M., and Hillenkamp, F. (1988) *Anal. Chem.* **60**, 2299-2301
153. Domon, B., and Aebersold, R. (2006) *Science* **312**, 212-217
154. Herbert, C. G., and Johnstone, R. A. W. *Mass Spectrometry Basics*, 1 Ed., CRC; 1st edition (June 26,2002)
155. Mann, M., and Jensen, O. N. (2003) *Nature Biotechnology* **21**, 255-261
156. Gevaert, K., and Vandenkerckhove, J. (2000) *Electrophoresis* **21**, 1145-1154
157. Yanisch-Perron, C., Viera, J., „ and Messing, J. (1985) *Gene* **33**, 103-108
158. Grant, S. G. N. (1990) *Proceedings of the National Academy of Sciences of the United States of America* **87**, 4645-4649
159. Phillips, T. A., Van Bogelen, R. A., and Neidhardt, F. C. (1984) *Journal of Bacteriology* **159**, 283-278
160. Wood, W. (1966) *J. Mol. Biol.* **16**, 118-133
161. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*
162. Sarker, G., and Sommers, S. S. (1990) *BioTechniques* **8**, 404-407
163. Laemmeli, U. K. (1970) *Nature* **227**, 680-685
164. Bui, B. T. S., Florentin, D., Fournier, F., Ploux, O., Mejean, A., and Marquet, A. (1998) *Febs Letters* **440**, 226-230

165. Sarker, G., and Sommers, S. S. (1990) *BioTechniques* **8**, 404-407
166. Bradford, M. M. (1976) *Anal. Biochem.* **72**
167. Konnermann, L., and Douglas, D. L. (1998) *Rapid Commum Mass Spectrometry* **12**, 435-442
168. Konnermann, L., Collings, B. A., and Douglas, D. L. (1997) *Biochemistry* **36**, 5554-5559
169. Konnermann, L., and Douglas, D. L. (1998) *Journal of Amercian Society of Mass Spectrometry* **9**, 1248-1254
170. Lee, V. W., Chen, Y. L., and Konnermann, L. (1999) *Anal. Chem.* **71**, 4154-4159
171. Konnermann, L., Rosell, F. I., Mauk, A. G., and Douglas, D. L. (1997) *Biochemistry* **36**, 6448-6456
172. Leon, S., Touraine, B., Brait, J.-F., and Lobreaux, S. (2002) *Biochem. J.* **366**, 557-564
173. Hernandez, H., Hewitson, K. S., Roach, P., Shaw, N. M., Baldwin, J. E., and Robinson, C. V. (2001) *Anal Chem* **73**, 4154-4161
174. Bui, B. T. S., Benda, R., Schunemann, V., Florentin, D., Trautwein, A. X., and Marquet, A. (2003) *Biochemistry* **42**, 8791-8798

Appendix


E.coli
A.aeolicus
A.thaliana
B.sphaericus
B.subtilis
B.flavum
C.pnuemoniae
E.herbicola
H.influenzea
H.pylori
H.pylori
M.autotrophicum
M.jannaschii
Methylobacillus
M.leprae
M.tuberculosis
P.stutzeri
P.abyssi
S.cerevisiae
S.pombe
S.pombe
S.marcescens
Synychocystis
T.maritima
Conservation
Mutation


230
240
250
260
270
280
290
300
310
320
330
340

Sequence and secondary structure of 6HisBS